Risk sensitive identification of linear stochastic systems *

László Gerencsér † György Michaletzky ‡
gerencser@sztaki.hu michgy@ludens.elte.hu

Zsuzsanna Vágó §
vago@sztaki.hu

Abstract

Risk-sensitive identification of AR-processes was first considered in [12]. The purpose of this paper is to extend this original approach to ARMA-processes and even multi-variable linear stochastic systems. We provide a new definition of a risk-sensitive identification criterion. For this we first consider a recursive identification procedure which is parameterized by a weight-matrix K acting on the stochastic gradient. Using the asymptotic theory of recursive estimation a suitably scaled version of the error process will be approximated by a stationary Gaussian process, see Chapter 4.5, Part II of [1]. The new risk sensitive criterion will be defined in terms of this associated stationary Gaussian process in a familiar manner via an exponential-quadratic cost. The main result of the paper is the minimization of the proposed new criterion with respect to the weight-matrix K over a feasible set E_{K} where the cost function is known to be finite, Theorem 6.1. This results will then be extended to the case when minimization

*This research was supported in part by grants from the Swedish Research Council for Engineering Sciences (TFR), the Göran Gustafsson Foundation, the National Research Foundation of Hungary (OTKA) under Grants T015668, T16665, T 020984 and T032932.

†MTA SZTAKI (Computer and Automation Institute of the Hungarian Academy of Sciences), H-1111 Kende u. 13-17, Budapest, Hungary

‡Department of Probability Theory and Statistics, Eötvös Loránd University, H-1117 Pázmány Péter sétány 1/C, Budapest, Hungary, and MTA SZTAKI (Computer and Automation Institute of the Hungarian Academy of Sciences), H-1111 Kende u. 13-17, Budapest, Hungary

§MTA SZTAKI (Computer and Automation Institute of the Hungarian Academy of Sciences), H-1111 Kende 13-17, Budapest, Hungary, and Pázmány Péter Catholic University, H-1088 Szentkirályi u. 28, Budapest, Hungary

over a feasible set E_{K}° is considered, on the complement of which the cost function is known to be infinite, Theorem 6.1. The starting point of our analysis is an expression of the cost function given in LEQG-theory (see [5]), in particular a result of [10]. A new expression for the cost function will be also given, using stochastic realization theory, as the mutual information rate between two stochastic processes.
Keywords: ARMA processes; linear stochastic systems; recursive prediction error identification; risk sensitive identification; Riccati equations; stochastic realization; bounded-real lemma.

1 Risk-sensitive identification of ARMA processes

Risk-sensitive identification of Gaussian AR-processes was first considered in [12], see pp. 297, Problem 3.2 of [12]. Let $(y_n),\ -\infty < n < +\infty$ be a (strictly) stationary Gaussian AR (p) process satisfying the difference equation

$$A^* y = e,$$

where A^* is polynomials of the backward shift operator of degree p. We assume that A^* is stable and the leading coefficients of A^* is equal to 1. The remaining coefficients of A^* are collected in a parameter vector θ^*. The noise process $e = (e_n)$ is assumed to be an independent, identically distributed Gaussian $\mathcal{N}(0, \sigma^2)$ sequence. Let

$$\mathcal{F}_n = \mathcal{F}_n^y = \sigma\{y_i : 0 \leq i \leq n\}$$

be the σ-algebra defined by the history of y between 0 and n. Then a special case of Problem 3.2 of [12] can be stated as follows: for a fixed time horizon N find a sequence of estimates $\widehat{\theta}_n,\ n = 1, ..., N$ such that for each n the estimate $\widehat{\theta}_n$ is $\mathcal{F}_n$-measurable and the criterion

$$\mathrm{E}\left(\exp\left\{ \frac{c}{2} \sum_{n=1}^{N} (\theta - \widehat{\theta}_n)^T (\theta - \widehat{\theta}_n) \right\} \right) , \tag{1.1}$$

with a positive c, is minimized. A solution to this problem is proposed in [12]. It is stated that, under additional technical conditions, the optimal solution is given by a recursion

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{1}{n} K_n (y_n - \phi_n^T \widehat{\theta}_{n-1}), \tag{1.2}$$

where K_n is a time-varying weighting matrix that can be computed by a forward recursion, and $\phi_n = (-y_{n-1}, ..., -y_{n-p})^T$.

The purpose of this paper is to extend these ideas by formulating a similar, but new risk sensitive identification criterion, which is applicable to ARMA-processes and even to multi-variable linear stochastic systems. Consider a piecewise constant embedding defined by

$$\widehat{\theta}_s = \widehat{\theta}_n \quad \text{for} \quad s \in [n, n+1), \quad n \geq 1.$$

Introducing the normalized and re-scaled process

$$\psi_t = e^{t/2}(\widehat{\theta}_{e^t} - \theta^*), \quad t \geq 0 \tag{1.3}$$

and $N = e^T$, the criterion given by (1.1) will become

$$\mathrm{E}\left(\exp\{\frac{c}{2}\int_0^T \psi(t)^T \psi(t) dt\} \right). \tag{1.4}$$

If $\widehat{\theta}_n$ is defined by a recursive procedure of the form (1.2) with fixed $K = K_n$, then, under appropriate technical conditions the asymptotic theory developed in Chapter 4.5, Part II of [1] is applicable, and we can approximate (ψ_t) by a stationary Gaussian-process. This observation is the basis of extending the ideas of [12].

Let us now consider the case of ARMA systems. Let $(y_n), \; -\infty < n < +\infty$ be a wide-sense stationary ARMA (p, q) process satisfying the difference equation

$$A^* y = C^* e,$$

where A^* and C^* are polynomials of the shift operator of degree p and q respectively. We assume, that A^*, C^* are stable, relative prime, and the leading coefficients of A^* and C^* are equal to 1. The remaining coefficients of A^* and C^* are collected in a parameter vector θ^*. At this point it is sufficient to assume that the noise process fulfills the following minimal conditions: there exists an increasing family of σ-algebras $(\mathcal{F}_n)$ such that e_n is $\mathcal{F}_n$-measurable and

$$\mathrm{E}(e_n|\mathcal{F}_{n-1}) = 0, \quad \mathrm{E}(e_n^2|\mathcal{F}_{n-1}) = \sigma^2 = \text{const}.$$

Under the above conditions we can proceed as follows. Let $D_\Theta \subset \mathbb{R}^{p+q}$ denote the set of system parameters such that the corresponding polynomials A and C are stable. For fixed $\theta \in D_\Theta$ define the process $\overline{\varepsilon}(\theta) = (\overline{\varepsilon}_n(\theta))$ by the difference equation $C\overline{\varepsilon} = Ay$, i.e.

$$\overline{\varepsilon} = (A/C)(C^*/A^*)e,$$

with $\overline{\varepsilon}_n = y_n = 0$ for $n \leq 0$. The asymptotic cost function is defined by

$$W(\theta) = \lim_{n\to\infty} \frac{1}{2\sigma^2} \mathrm{E}\overline{\varepsilon}_n^2(\theta).$$

It is well known and easily shown, that

$$\frac{\partial}{\partial\theta} W(\theta)\bigg|_{\theta=\theta^*} = 0 \quad \text{and} \quad R^* \stackrel{\Delta}{=} \frac{\partial^2}{\partial\theta^2} W(\theta)\bigg|_{\theta=\theta^*} > 0.$$

To define a weighted recursive prediction error identification method (cf. [11]) let $\widehat{\theta}_n$ denote the estimator of θ^* at time n and let the on-line estimate of $\overline{\varepsilon}_n(\widehat{\theta}_{n-1})$ be denoted by ε_n. They are constructed as follows. Let $\widehat{\theta}_0 \in D_\Theta$ be an arbitrary initial guess and set $\varepsilon_n = y_n = 0$ for $n \leq 0$. Now, if $\widehat{\theta}_k$ and ε_k have already been generated for $k \leq n-1$ then define ε_n by the equation:

$$\left(\widehat{C}_{n-1}\varepsilon\right)_n = \left(\widehat{A}_{n-1}y\right)_n, \tag{1.5}$$

where $\widehat{A}_{n-1}$, $\widehat{C}_{n-1}$ denote the polynomials corresponding to $\widehat{\theta}_{n-1}$ and $(\cdot)_n$ denotes evaluation at time n. Similarly, we define the on-line estimate of the gradient of the process $\overline{\varepsilon}(\theta)$, denoted by ε_θ, by

$$\left(\widehat{C}_{n-1}\varepsilon_\theta\right)_n = -\phi_{n-1}, \tag{1.6}$$

where $\phi_{n-1} = (-y_{n-1}, \cdots - y_{n-p}, \varepsilon_{n-1}, \ldots \varepsilon_{n-q})^T$. Then the weighted recursive prediction error estimate of θ^* at time N is defined by the recursion

$$\widehat{\theta}_n = \widehat{\theta}_{n-1} - \frac{1}{n} K \varepsilon_{\theta N} \varepsilon_n, \tag{1.7}$$

where K is a *fixed* weighting matrix.

The asymptotic properties of $\widehat{\theta}_n$ have been rigorously analyzed in [1] and [3] under various technical conditions. In [1] it is required that $\hat{\theta}_n \in D_0 \subset D_\Theta$, where D_0 is a prescribed compact domain, otherwise the process is stopped. In [3] the boundedness condition above is enforced by a resetting mechanism: if $\hat{\theta}_n \notin D_0$ then we redefine it to be $\widehat{\theta}_1$ again. To sketch a key result of [1] define the estimation sequence $\widehat{\theta}_{n,k}$ using the recursion (1.7) but changing the step-sizes from $1/N$ to $1/(N+k)$ and consider a piecewise constant embedding defined by

$$\widehat{\theta}_{k,s} = \widehat{\theta}_{k,n} \quad \text{for} \quad s \in [n, n+1), \quad n \geq 1.$$

Introduce a normalized and re-scaled process by first normalizing $\widehat{\theta}_{k,s} - \theta^*$ by $s^{1/2}$, followed by an exponential change of time-scale $s = e^t$. This yields a new process $\psi_k = (\psi_{k,t})$:

$$\psi_{k,t} = e^{t/2}(\widehat{\theta}_{k,e^t} - \theta^*), \quad t \geq 0. \tag{1.8}$$

It is claimed in Theorem 12 Chapter 4.5, Part II of [1] that under appropriate technical conditions $(\psi_{k,t})$ converges weakly, for $k \to \infty$, to the process $(\widetilde{x}(t))$ defined by

$$d\widetilde{x}(t) = (-KR^* + I/2)\widetilde{x}(t)dt + Gdw(t), \tag{1.9}$$

where $(w(t))$ is a standard $\mathbb{R}^{p+q}$-valued Wiener-process and G is the symmetric positive semi-definite square-root of KR^*K^T, i.e. $G = G^T$ and

$$GG = KR^*K^T,$$

assuming that

$$F = -KR^* + I/2$$

is asymptotically stable, i.e. all the eigenvalues of F are on the open left half of the complex plane.

For the recursive estimation procedure with enforced boundedness, that has been rigorously analyzed in [3], a corresponding result has not yet been fully derived, but a significant part of the relevant analysis has been completed in [4].

A direct corollary of the cited result of [1] is that the asymptotic covariance matrix $S = S(K)$ of the estimator process $\widehat{\theta}_n$ exists. Obviously, it is given by

$$S = \mathrm{E}\widetilde{x}(t)\widetilde{x}(t)^T ,$$

and it is well-known that S is the solution of the Lyapunov-equation

$$(-KR^* + I/2)S + S(-KR^* + I/2)^T + KR^*K^T = 0. \tag{1.10}$$

Using the partial ordering for symmetric matrices: $A \leq B$ if and only if $B - A$ is positive semidefinite, where A, B are symmetric matrices, it is also well-known that $S = S(K)$ is minimized with respect this ordering for the choice $K = (R^*)^{-1}$. Then $F = -I/2$ and for the asymptotic covariance matrix of the error-process we have $S = (R^*)^{-1}$.

A risk-sensitive criterion for the identification of ARMA-processes will be defined along the lines of (1.1) and (1.4) by

$$J(K) = \frac{2}{c} \lim_{T\to\infty} \frac{1}{T} \log \mathrm{E} \left(\exp\{\frac{c}{2} \int_0^T \widetilde{x}(t)^T H^T H \widetilde{x}(t) dt\} \right), \tag{1.11}$$

where $c > 0$ and $H^T H$ is a non-singular weighting-matrix. The minimization of this criterion with respect to K is the subject matter of this paper. Two domains of definitions of $J(K)$ will be considered. First we consider a feasible set E_{K}, defined in (3.3), where the cost function is known to be finite, see Corollary 6.1. Then we consider a feasible set E_{K}°, defined in (3.7), on the complement of which the cost function is known to be infinite, see Corollary 7.1.

!!! Is the closure of E_{K} E_{K}°?

2 Multi-variable linear stochastic systems

The extension of the above setup to multi-variable linear stochastic systems is obtained as follows: Consider the state-space system

$$\left\{ \begin{array}{rcl} \zeta_{n+1} & = & A(\theta)\zeta_n + B(\theta)e_n \\ y_n & = & C(\theta)\zeta_n + e_n \end{array} \right. \tag{2.1}$$

with $-\infty < n < +\infty$, where the function

$$H(\theta) = H(z,\theta) = I + C(\theta)\left(zI - A(\theta)\right)^{-1} B(\theta) \ ,$$

is an $m \times m$ square transfer-function. For the sake of convenience we consider the processes above defined in $-\infty < n < +\infty$. Thus the process $y = (y_n)$ is interpreted as the unique, weakly stationary solution of (2.1). Here θ is a parameter vector belonging to an open domain $D_\Theta \subset \mathbb{R}^p$.

Condition 2.1 *The transfer functions $H(\theta)$, $\theta \in D_\Theta \subset \mathbb{R}^p$, are stable and inverse-stable. Moreover $A(\theta), B(\theta), C(\theta)$ are twice continuously differentiable with respect to θ.*

Condition 2.2 *There exists an increasing family of σ-algebras $(\mathcal{F}_n)$ such that e_n is $\mathcal{F}_n$-measurable and*

$$\mathrm{E}(e_n|\mathcal{F}_{n-1}) = 0, \quad \mathrm{E}\left(e_n e_n^T\right)|\mathcal{F}_{n-1}) = \Lambda^* \ , \quad \text{i.e. it is constant for all } n,$$

where

$$\Lambda^* > 0.$$

It follows that (e_n) is the innovation process of (y_n).

The set of symmetric, positive definite $m \times m$ matrices is denoted by D_Λ. Define for $\theta \in D_\Theta$ the estimated innovation process by

$$\bar{\varepsilon}(\theta) = H(\theta)^{-1} y \ .$$

Then the asymptotic cost function is defined for $\theta \in D_\Theta$, $\Lambda \in D_\Lambda$ by

$$W(\theta,\Lambda) = \frac{1}{2} E\left(\bar{\varepsilon}_n^T(\theta)\Lambda^{-1}\bar{\varepsilon}_n(\theta)\right) + \frac{1}{2}\log\det\Lambda \ .$$

Note that if (e_n) is an i.i.d. sequence of Gaussian random vectors with distribution $N(0,\Lambda^*)$, then $W(\theta,\Lambda)$ is the asymptotic negative log-likelihood function, except for an additive constant. The gradient of $W(\theta,\Lambda)$ with respect to θ and Λ^{-1} is given by

$$\begin{array}{rcl} W_\theta(\theta,\Lambda) & = & \mathrm{E}\ \bar{\varepsilon}_{\theta n}^T(\theta)\Lambda^{-1}\bar{\varepsilon}_n(\theta) \\ W_{\Lambda^{-1}}(\theta,\Lambda) & = & \dfrac{1}{2}\left(\mathrm{E}\ \bar{\varepsilon}_n(\theta)\bar{\varepsilon}_n^T(\theta) - \Lambda\right). \end{array}$$

Here the gradients of the components of $\bar{\varepsilon}_n(\theta)$ are represented as column-vectors. Set

$$R_1^* = W_{\theta\theta}(\theta^*, \Lambda^*) = \mathrm{E}\ \bar{\varepsilon}_{\theta n}^T(\theta^*)\Lambda^{*-1}\bar{\varepsilon}_{\theta n}(\theta^*)\ . \tag{2.2}$$

Then the Hessian of $W(\theta, \Lambda)$ at (θ^*, Λ^*) is

$$R^* = \begin{pmatrix} R_1^* & 0 \\ 0 & \Lambda^* \otimes \Lambda^* \end{pmatrix}. \tag{2.3}$$

Condition 2.3 *We assume that for any fixed* $\Lambda > 0$ *the equation*

$$W_\theta(\theta, \Lambda) = 0$$

has a unique solution $\theta = \theta^*$ *in* D_Θ*, and*

$$R_1^* = W_{\theta\theta}(\theta^*, \Lambda^*) > 0\ .$$

A weighted recursive prediction estimation of (θ^*, Λ^*) is obtained as follows. First define the correction terms

$$\begin{aligned} H_{1n}(\theta, \Lambda^{-1}) &= \bar{\varepsilon}_{\theta n}^T(\theta)\Lambda^{-1}\bar{\varepsilon}_n(\theta) \\ H_{2n}(\theta, \Lambda^{-1}) &= \frac{1}{2}(\ \bar{\varepsilon}_n(\theta)\bar{\varepsilon}_n^T(\theta) - \Lambda) \end{aligned}$$

and then consider the recursion

$$\begin{bmatrix} \theta_n \\ \Lambda_n^{-1} \end{bmatrix} = \begin{bmatrix} \theta_{n-1} \\ \Lambda_{n-1}^{-1} \end{bmatrix} - \frac{1}{n}K \begin{bmatrix} H_{1n}(\theta_{n-1}, \Lambda_{n-1}^{-1}) \\ H_{2n}(\theta_{n-1}, \Lambda_{n-1}^{-1}) \end{bmatrix}\ .$$

The recursion above is a frozen-parameter recursion from which a genuinely recursive estimation is obtained in a standard way (cf. [1] or [8]). Set

$$H_n = (H_{1n}, H_{2n}).$$

It is easily seen that at $(\theta, \Lambda) = (\theta^*, \Lambda^*)$ the sample covariance-matrix of the process H_n is given by

$$S^* = \begin{pmatrix} R_1^* & 0 \\ 0 & \frac{1}{4}\ \mathrm{E}\ (ee^T \otimes ee^T - \Lambda^* \otimes \Lambda^*) \end{pmatrix}\ . \tag{2.4}$$

Omitting technical details and conditions we note that just as in the ARMA-case, using a piecewise constant embedding, the normalized, re-scaled and re-initialized estimation error process converges weakly to a $r = p + m^2$-dimensional process $(\widetilde{x}(t))$ defined by

$$d\tilde{x}(t) = \left(-KR^* + \frac{I}{2}\right)\tilde{x}(t)dt + Gdw(t)\ , \tag{2.5}$$

where $(w(t))$ is a standard $\mathbb{R}^{p+m^2}$-valued Wiener-process and G is the symmetric positive semi-definite square-root of KS^*K^T, i.e. $G = G^T$ and

$$GG = KS^*K^T,$$

assuming that $F = -KR^* + I/2$ is asymptotically stable. Equation (2.5) looks very much the same as equation (1.9) for the ARMA case, but there is a major difference: the covariance matrix KS^*K^T is obtained here instead of KR^*K^T and generically

$$S^* \neq R^*.$$

A risk-sensitive identification criterion will be defined by

$$J(K) = \frac{2}{c} \lim_{T\to\infty} \frac{1}{T} \log \mathrm{E} \left(\exp \left\{ \frac{c}{2} \int_0^T \widetilde{x}(t)^T H^T H \widetilde{x}(t)\, dt \right\} \right) , \tag{2.6}$$

where $c > 0$, and H is a non-singular $r \times r$ matrix. Note that this is a mixed criterion in the sense that $J(K)$ is defined in terms of both $\hat{\theta}_n$ and $\hat{\Lambda}_n$. The domain of definition of J will be made clear in below, see (3.3) and (3.7). An explicit solution to the problems of minimizing $J(K)$ will be given in the special case when the H is block-diagonal of appropriate dimensions, see Corollary 6.2.

A *general* risk-sensitive optimization problem can now be formulated as follows: for given matrices R^*, S^*, where R^* is an $r \times r$, symmetric, positive semi-definite matrix, and S^* is an $r \times r$, symmetric, positive semi-definite matrix

!!! definite or semi-definite ???

consider the state-space equation, with $\tilde{x}(t) \in \mathbb{R}^r$,

$$d\tilde{x}(t) = \left(-KR^* + \frac{I}{2} \right) \tilde{x}(t)dt + Gdw(t) , \tag{2.7}$$

where K is an $r \times r$ matrix,

$$F = -KR^* + I/2$$

is asymptotically stable, and G is the unique symmetric positive semi-definite square-root of KS^*K^T, thus in particular

$$GG = KS^*K^T,$$

finally $(w(t))$ is a standard Wiener-process in $\mathbb{R}^r$. Then to find a K for which the function $J(K)$

$$J(K) = \frac{2}{c} \lim_{T\to\infty} \frac{1}{T} \log \mathrm{E} \left(\exp \left\{ \frac{c}{2} \int_0^T \widetilde{x}(t)^T H^T H \widetilde{x}(t)\, dt \right\} \right) , \tag{2.8}$$

where $c > 0$, and H is a non-singular $r \times r$ matrix, is minimized. The criterion (2.8) is the continuous-time version of the functional defined in [12] (see (3.2) of [12]).

The main result of the paper is the minimization of the proposed new criterion with respect to the weight-matrix K over a feasible set E_{K}°, defined in (3.7), on the complement of which the cost function is known to be infinite, see Theorem 7.1. This general result is obtained in several steps. The starting point of our analysis is an expression of the cost function given in LEQG-theory, in particular a result of [10]. First we determine the unique stationary point of $J(K)$ over a feasible set E_{K}, defined in (3.3), where the cost function is known to be finite, see Theorem 5.1. The stationary point of $J(K)$ will be found by solving an extended, and relaxed matrix-valued constrained minimization problem, where the equality constraints are defined by a control-Riccati equation. Then, using a filter-Riccati representation of $J(K)$, we verify that this stationary point is in fact the unique minimum of $J(K)$ over E_{K}, see Theorem 6.1. Finally, by letting the parameter c vary we derive the main result.

A new expression for the cost function will be also given as the mutual information rate between two stochastic processes. These processes are constructed from the normalized and re-scaled estimation error process by using stochastic realization theory. The interpretation of our main results on minimizing $J(K)$ in this context is, however, still to be explored.

3 Expressing $J(K)$ via Riccati equations

Functionals of the form of $J(K)$ are well-known in risk-sensitive control or LEQG (linear exponential quadratic gaussian) control and a number of useful expressions for $J(K)$ have been found. LEQG control has been first defined by Jacobson in 1973, see [5], followed by various extensions in Speyer, Deyst and Jacobson, [13], Kumar, [6], Kumar and Van Schuppen, [7], Whittle [14], and by Bensoussan and Van Schuppen, [2]. Two good surveys of the area are [10] and [12]. The following proposition is given as Proposition 6.3.1 in [10]. Define the transfer function $\mathcal{G}$ by

$$\mathcal{G} = H(sI - F)^{-1}G. \tag{3.1}$$

Proposition 3.1 *Assume that $F = -KR^* + I/2$ is asymptotically stable. Then:*

(i) If the H_∞-norm of $c^{1/2}\mathcal{G}(s)$ is strictly less than 1, then $J(K)$ is well-defined and finite and

$$J(K) = \lim_{s_0 \to \infty} -\frac{1}{2\pi c} \int_{-\infty}^{\infty} \ln|\det(I - c\mathcal{G}(i\omega)\mathcal{G}^*(i\omega))| \left[\frac{s_0^2}{s_0^2 + \omega^2}\right] d\omega. \tag{3.2}$$

(ii) If the H_∞-norm of $c^{1/2}\mathcal{G}(s)$ is greater than 1, then $J(K)$ is well-defined and $J(K) = \infty$.

A similar result has been established for discrete time linear stochastic Gaussian systems in [12].

We will use the following notations: the set of K-s for which $F = -KR^* + I/2$ is asymptotically stable will be denoted by D_{K}:

$$D_{\mathrm{K}} = \{K \in \mathbb{R}^{r\times r} : \ F = -KR^* + I/2 \text{ is asymptotically stable}\}.$$

Obviously D_{K} is an open domain in the set of $r \times r$ matrices. Next consider the set of K-s for which the H_∞-norm of $c^{1/2}\mathcal{G}$ defined in (3.1) is strictly less than 1:

$$E_{\mathrm{K}} = \{K : \ K \in D_{\mathrm{K}} \text{ and } \|c^{1/2}\mathcal{G}\|_\infty < 1\}. \tag{3.3}$$

According to (i) of the previous proposition $J(K)$ is well-defined and finite on E_{K}. A computationally useful expression for $J(K)$ on E_{K} can be given using Lemma 5 and Theorem 5 of [15] and Proposition 5.3.2 of [10]:

Proposition 3.2 *Assume that $F = -KR^* + I/2$ is asymptotically stable. Then*

(i) the H_∞-norm of $c^{1/2}\mathcal{G}(s)$ is less than or equal to 1 if and only if the control-Riccati equation:

$$F^TQ + QF + H^TH + cQKS^*K^TQ = 0 \tag{3.4}$$

has a real symmetric solution Q for which the matrix $F + cKS^*K^TQ$ is stable. Moreover, this solution is unique and positive definite.

(ii) the H_∞-norm of $c^{1/2}\mathcal{G}(s)$ is strictly less then 1 if and only if the control-Riccati equation above has a real symmetric solution, for which the matrix $F+cKS^*K^TQ$ is asymptotically stable.

Furthermore, in case (ii) the functional $J(K)$ is finite and

$$J(K) = \mathrm{tr}\ QKS^*K^T = \mathrm{tr}\ GQG. \tag{3.5}$$

A real, symmetric solution Q of (3.4) for which $F + cKS^*K^TQ$ is asymptotically stable will be called a stabilizing solution. Proposition 3.1 implies that on the set

$$\{K : \ K \in D_{\mathrm{K}} \text{ and } \|c^{1/2}\mathcal{G}\|_\infty > 1\}$$

$J(K)$ is well-defined and $J(K) = \infty$.

In the critical case $\|c^{1/2}\mathcal{G}\|_\infty = 1$ we are unaware of any definite result on the existence of a well-defined $J(K)$. We will therefore extend the definition of $J(K)$ by writing

$$J(K) = \frac{2}{c} \liminf_{T\to\infty} \frac{1}{T} \log \mathrm{E}\left(\exp\{\frac{c}{2} \int_0^T \widetilde{x}(t)^T H^T H \widetilde{x}(t) dt\} \right). \tag{3.6}$$

Also define an extension of E_{K} as follows:

$$E_{\mathrm{K}}^{\circ} = \{K : \ K \in D_{\mathrm{K}} \ , \|c^{1/2}\mathcal{G}\|_{\infty} \leq 1\} \ . \tag{3.7}$$

Obviously $E_{\mathrm{K}} \subset E_{\mathrm{K}}^{\circ}$.

!!! Is the closure of E_{K} E_{K}° ?

The minimization of $J(K)$ over E_{K}° will considered in Section 7. The next proposition gives a simple necessary condition for E_{K}° not being empty.

Proposition 3.3 *Assume that there exists a matrix K, for which $-KR^* + \frac{I}{2}$ asymptotically stable and $\|c^{1/2}\mathcal{G}\|_{\infty} \leq 1$. Then*

$$S^* > cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*.$$

Proof. According to Proposition 3.2 there exists a symmetric, positive definite solution Q of the Riccati equation (3.4). Substituting $F = -KR^* + I/2$ into (3.4), multiplying by c and by $S^*(R^*)^{-1}$ from the left, $(R^*)^{-1}S^*$ from the right and completing to squares the terms containing K, we arrive at the equation

$$\begin{aligned} cS^*(R^*)^{-1}Q(R^*)^{-1}S^* \quad &+ \quad \left(cS^*(R^*)^{-1}QK - I\right) S^* \left(cK^T Q(R^*)^{-1}S^* - I\right) \\ &= \quad S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*, \end{aligned}$$

implying that $S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*$ is positive definite.

Remark. The above necessary condition is also sufficient for E_{K}° and even E_{K} not being empty, see Theorem 5.1.

Remark. Note that in the case when $S^* = R^*$, such as the case of ARMA-systems, the necessary condition reduces to

$$R^* > cH^T H.$$

An alternative expression for $J(K)$ is given in terms of a filter-Riccati equation in the following proposition, which is implied by Proposition 2.3.1 of [10] or Lemma 8 of [15]:

Proposition 3.4 *Assume that $F = -KR^* + I/2$ is asymptotically stable. Then the H_{∞}-norm of $c^{1/2}\mathcal{G}$ is strictly less than 1 if and only if the filter-Riccati equation*

$$FP + PF^T + cPH^T HP + GG = 0 \tag{3.8}$$

has a real, symmetric solution, for which the matrix $F + cPH^T H$ is asymptotically stable. This solution is unique and positive definite. In this case the functional $J(K)$ is well-defined, finite and can be written as

$$J(K) = \mathrm{tr}\ PH^T H. \tag{3.9}$$

Recall that if $F + cPH^T H$ is asymptotically stable then P is called a stabilizing solution.
Remark. It follows immediately that, for sufficiently small c, $J(K)$ is finite and for $c \searrow 0$ $J(K)$ converges to

$$\mathrm{tr}\ S(K)H^T H = \mathrm{E}\widetilde{x}^T(s)H^T H\widetilde{x}(s),$$

where $S(K)$ is the solution of the Lyapunov-equation

$$FS + SF^T + GG = 0.$$

4 $J(K)$ as a mutual information rate

In [12] Stoorvogel and Van Schuppen introduced a number of information theoretic criteria for system identification in the case of discrete-time stochastic processes. They prove among others that the mutual information rate of the error process of a parameter estimation obtained from a recursive estimation scheme and an appropriately defined white noise coincides with the LEQG functional. Proposition 4.1 below claims that such a representation is possible in the present continuous-time case, as well. Note that the proof in [12] is based on an application of Szegő's theorem, which not applicable in continuous-time.

Consider the state-space equation

$$dx(t) = Fx(t)dt + c^{1/2}Gdw(t), \tag{4.1}$$

where $w(t)$ is a standard Wiener process. Note that $x(t) = c^{1/2}\widetilde{x}(t)$. Extend this system with

$$d\xi(t) = F\xi(t)dt - cPH^T db(t) \tag{4.2}$$

where $b(t)$ is a standard Wiener process, independent of $w(t)$ and $\xi(t)$ is the stationary solution of (4.2).

Then the process y defined by

$$dy(t) = H(x(t) + \xi(t))dt + db(t) \tag{4.3}$$

is a standard Wiener process, i.e. the transfer function

$$\mathcal{N}(s) = I - H(sI - F)^{-1}cPH^T$$

provides a so-called all-pass "extension" of the transfer function $c^{1/2}\mathcal{G}(s)$. In other words $\left[c^{1/2}\mathcal{G}(s), \mathcal{N}(s)\right]$ is all-pass (in particular it is inner, due to the stability of F), mapping the Wiener process $\begin{bmatrix} w(t) \\ b(t) \end{bmatrix}$ into the Wiener process $y(t)$. Indeed, direct computation shows that a factorization of $I - c\mathcal{G}\mathcal{G}^*$ is given by $\mathcal{N}\mathcal{N}^*$, i.e.

$$I - c\mathcal{G}(i\omega)\mathcal{G}^*(i\omega) = \mathcal{N}(i\omega)\mathcal{N}^*(i\omega).$$

(Cf. Mustafa and Glover [10] Lemma 5.3.2.)

It is easy to see that a minimal realization of $\left[c^{1/2}\mathcal{G}(s), \mathcal{N}(s)\right]$ is given by

$$\left\{ \begin{array}{rcl} d\,(x(t)+\xi(t)) & = & F\,(x(t)+\xi(t))\;dt + \left[c^{1/2}G,\; -cPH^T\right] \begin{bmatrix} w(t) \\ b(t) \end{bmatrix} \\ dy(t) & = & H\,(x(t)+\xi(t))\;dt + db(t) \end{array} \right. \tag{4.4}$$

Proposition 4.1 *Assume that the $\|c^{1/2}\mathcal{G}\|_\infty < 1$. Consider the auxiliary process y defined by (4.3) and let $I(y,w)$ denote the mutual information rate between the processes y and w. Then*

$$J(K) = \frac{2}{c} I(y,w) = \text{tr } PH^T H\ . \tag{4.5}$$

Proof. Recall that the mutual information rate between two processes is defined as follows. Consider a finite value T and denote by $Q_{y,T}$ and $Q_{w,T}$ the distribution of the processes $y(s), 0 \le s \le T$ and $w(s), 0 \le s \le T$ defined on the space of continuous vector-valued functions and let $Q_{y,w,T}$ denote their joint distribution. Then

$$I(y,w) = -\lim_{T\to\infty} \frac{1}{T} E\left(\ln \frac{d\,(Q_{y,T}\times Q_{w,T})}{dQ_{y,w,T}}\,(y(s),w(s),0\le s\le T)\right)\ , \tag{4.6}$$

assuming that the limit exists.

The quantity $E\left(-\ln \frac{d\left(Q_{y,T}\times Q_{w,T}\right)}{dQ_{y,w,T}}\,(y(s),w(s),0\le s\le T)\right)$ is the divergence between the joint distribution of y, w and the product measure with the marginal distributions given by y and w separately – taken in the time interval $[0,T]$.

Let us write equations (4.1), (4.2) and (4.3) in the following form

$$\begin{array}{rcl} d(x(t)+\xi(t)) & = & F(x(t)+\xi(t))dt + \left[c^{1/2}G, -cPH^T\right] \begin{bmatrix} dw(t) \\ db(b) \end{bmatrix} \\ d\begin{bmatrix} w(t) \\ y(t) \end{bmatrix} & = & \begin{bmatrix} 0 \\ H \end{bmatrix} (x(t)+\xi(t))dt + \begin{bmatrix} dw(t) \\ db(b) \end{bmatrix}\ . \end{array}$$

In the present case the process y is a Wiener-process, as we have already pointed out, consequently the product measure with marginal distributions defined by the processes w and y coincide with the joint distribution of w, b taken on the same time interval.

Thus Girsanov's theorem allows us to compute the Radon-Nikodym derivate evaluated at the values of y, w. Introducing the notation $\zeta = x + \xi$, we have

$$\frac{dQ_{y,T} \times dQ_{w,T}}{dQ_{y,w,T}} (y(s), w(s), 0 \leq s \leq T)$$
$$= \exp\left\{ - \int_0^T \zeta(t)^T \left[0, H^T\right] d \begin{bmatrix} dw(t) \\ db(b) \end{bmatrix} - \frac{1}{2} \int_0^T \zeta^T(t) \left[0, H^T\right] \begin{bmatrix} 0 \\ H \end{bmatrix} \zeta(t) dt \right\}.$$

Taking the logarithm and using that the first term is a square-integrable martingale with zero expectation we get that

$$I(y, w) = \lim_{T \to \infty} \frac{1}{T} E \left(\frac{1}{2} \int_0^T \zeta^T(t) \left[0, H^T\right] \begin{bmatrix} 0 \\ H \end{bmatrix} \zeta(t) dt \right) .$$

The covariance matrix of the stationary process ζ is cP, and thus

$$I(y, w) = \frac{c}{2} \mathrm{tr}\ P H^T H ,$$

proving that the mutual information rate between y and w exists. Taking into account the equation (3.9) the proposition follows.

Thus the minimization of $J(K)$ is equivalent to minimizing the mutual information rate between the fixed Wiener-process w and the K-dependent Wiener-process y.

5 Stationary points of $J(K)$ over E_{K}: a control-Riccati approach

In this section we prove that $J(K)$ has a unique stationary point in E_{K}, and determine its value, using the control-Riccati representation of $J(K)$. This is the content of Theorem 5.1. The fact that this stationary point is indeed the unique minimum of $J(K)$ over E_{K} will be verified in the next section. Since the derivation of the next section self-contained and is related to the present section only in using an intelligent guess for the optimal value of K, some of the details of the proof of Theorem 5.1 of the present section will be omitted.

The stationary point will be first determined for an extended, relaxed optimization problem, stated as a matrix-valued constrained minimization problem, where the equality constraints are defined by a control-Riccati equation, assuming that the solution is not on the boundary of the feasible set. It is then verified that the obtained stationary point is indeed a solution of the original problem.

Theorem 5.1 *Assume that* $S^* > cS^* R^{*-1} H^T H (R^*)^{-1} S^*$. *Then the set* E_{K} *defined under (3.3) is non-empty, and* $J(K)$ *has a unique stationary point in* E_{K} *given by*

$$K^* = \left(R^* - cS^*(R^*)^{-1}H^T H\right)^{-1} .$$

The corresponding cost is

$$J(K^*) = \mathrm{tr}\ \left(R^* \left(S^*\right)^{-1} R^* - cH^T H\right)^{-1} H^T H .$$

Note that the optimal K^* may be non-symmetric, which is an unusual feature of risk-sensitive identification.

To prove Theorem 5.1 we now define an extended problem. Let the set of positive definite symmetric matrices Q be denoted by D_{Q}. Consider the extended variable (K, Q) with $K \in D_{\mathrm{K}}$, $Q \in D_{\mathrm{Q}}$ and, motivated by Proposition 3.2, define a *relaxed* constrained minimization problem as follows.

$$\begin{array}{lll} \text{minimize} & \mathrm{tr}\ QKS^*K^T & (5.1) \\ \text{subject to} & K \in D_{\mathrm{K}}, \quad Q \in D_{\mathrm{Q}}, & \\ & F^T Q + QF + H^T H + cQKS^*K^T Q = 0. & (5.2) \end{array}$$

Note that due to the nonsingularity of $H^T H$ any solution Q of the equation (5.2) is nonsingular.

By virtue of Proposition 3.2, if we add the constraint that $F + cKS^*K^TQ$ is asymptotically stable, then the constrained optimization problem is equivalent to minimizing $J(K)$ over E_{K}. For the cost function (5.1) we introduce the notation

$$J(K, Q) = \mathrm{tr}\ QKS^*K^T .$$

Theorem 5.2 *Assume that* $S^* > cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*$. *Then the feasible set for the constrained minimization problem (5.1), (5.2) is non-empty, and there is a unique* (K^*, Q^*) *satisfying the first order necessary conditions of optimality, given by*

$$\begin{aligned} K^* &= (R^* - cS^*(R^*)^{-1}H^T H)^{-1}, \\ Q^* &= (R^* - cH^T H(R^*)^{-1}S^*)(R^*)^{-1}H^T H \\ &= H^T H - cH^T H(R^*)^{-1}S^*(R^*)^{-1}H^T H. \end{aligned} \tag{5.3}$$

The corresponding cost is

$$J(K^*, Q^*) = \mathrm{tr}\ H^T H(R^* S^{*-1} R^* - cH^T H)^{-1}.$$

Proof. The Lagrangian of the constrained optimization problem is as follows:

$$L(K,Q,\Lambda) = \text{tr }(GQG + \Lambda(F^TQ + QF + H^TH + cQGGQ)), \tag{5.4}$$

where the Lagrange-multipliers can be assumed to be represented by a symmetric matrix Λ, due to the symmetry of the constraint given by the control-Riccati equation (5.2).

Computing the partial derivates with respect to K, Q and Λ we arrive at the following equations:

$$2QKS^* - 2Q\Lambda R^* + 2cQ\Lambda QKS^* = 0\ , \tag{5.5}$$

$$GG + \Lambda F^T + F\Lambda + c\Lambda QGG + cGGQ\Lambda = 0. \tag{5.6}$$

and the Riccati equation (5.2).

Multiplying (5.5) by Q^{-1} from the left, and by K^T from the right we get that

$$KS^*K^T - \Lambda R^*K^T + c\Lambda QKS^*K^T = 0. \tag{5.7}$$

Using $F = -KR^* + I/2$ in (5.6) we obtain that

$$KS^*K^T + \Lambda - \Lambda R^*K^T - KR^*\Lambda + c\Lambda QKS^*K^T + cKS^*K^TQ\Lambda = 0.$$

Subtracting (5.7) + $(5.7)^T$ from the latter equation we get:

$$\Lambda - KS^*K^T = 0\ . \tag{5.8}$$

We prove, that if $K \in E_{\mathrm{K}}$ then K and henceforth Λ is nonsingular. Indeed, assume that $x^TK = 0$ for some $x \neq 0$. Then $x^TF = x^T(-KR^* + I/2) = x^T/2$, and thus $1/2$ is an eigenvalue of F^T, and F^T would not be asymptotically stable.

Substituting (5.8) into the first term of (5.7) we obtain that

$$\Lambda - \Lambda R^*K^T + c\Lambda QKS^*K^T = 0\ ,$$

and multiplying by Λ^{-1} we get:

$$I - (R^* - cQKS^*)\,K^T = 0\ . \tag{5.9}$$

Now let us consider the Riccati equation (5.2) with the substitution $F = -KR^* + \frac{I}{2}$:

$$Q - R^*K^TQ - QKR^* + H^TH + cQKS^*K^TQ = 0\ .$$

Now multiplying (5.9) by Q we get

$$Q - R^*K^TQ + cQKS^*K^TQ = 0$$

and using the latter equality we get

$$-QKR^* + H^T H = 0$$

from which we get

$$QK = H^T H (R^*)^{-1} \,. \tag{5.10}$$

Inserting into (5.9) we arrive at the conclusion that the first order necessary conditions of optimality given in (5.5) and (5.6) together with the Riccati equation (3.4) uniquely determine K by

$$K^{-1} = R^* - cS^*(R^*)^{-1} H^T H \,. \tag{5.11}$$

We get Λ and Q from (5.8) and (5.10), respectively. In summary, the only possible stationary point is

$$\begin{aligned}
K^* &= (R^* - cS^*(R^*)^{-1}H^T H)^{-1} \\
\Lambda^* &= (R^* - cS^*(R^*)^{-1}H^T H)^{-1} S^* (R^* - cH^T H(R^*)^{-1} S^*)^{-1} \,, \\
Q^* &= H^T H (R^*)^{-1} (R^* - cS^*(R^*)^{-1} H^T H) \\
&\quad = H^T H - cH^T H (R^*)^{-1} S^* (R^*)^{-1} H^T H.
\end{aligned}$$

The invertibility of $(R^* - cS^*(R^*)^{-1}H^T H)$ follows from the condition

$$S^* > cS^*(R^*)^{-1} H^T H (R^*)^{-1} S^*.$$

The above triplet is the only possible stationary point and using the above argument in a reverse direction it can be verified that it is indeed a stationary point. The associated cost is

$$\begin{aligned}
J(K^*, Q^*) &= \operatorname{tr} Q^* K^* S^* K^{*T} = \operatorname{tr} Q^* \Lambda^* \\
&= \operatorname{tr} H^T H (R^*)^{-1} S^* (R^* - cH^T H (R^*)^{-1} S^*)^{-1} \\
&= \operatorname{tr} H^T H (R^* S^{*-1} R^* - cH^T H)^{-1} \,,
\end{aligned}$$

and with this the proof of Theorem 5.2 is complete.

Proof of theorem 5.1. First assume that $K^\circ \in E_{\mathrm{K}}$ satisfies the first order necessary condition of optimality for the minimization of $J(K)$. Let now K vary in a small neighborhood of K°, and let $Q = Q(K)$ be the unique, real, symmetric, positive definite solution of the control-Riccati equation (3.4), for which the matrix $F + cTKS^*K^TQ$ is asymptotically stable. It is easy to see that $Q(K)$ is a smooth function of K in a small neighborhood of K°. Thus it follows, by elementary calculus, that K° satisfies the first order necessary condition of optimality for the minimization of $J(K)$ if and only if

(K°, Q°) satisfy the first order necessary conditions of optimality for the the constrained minimization problem (5.1), (5.2). But then we must have $K^\circ = K^*, Q^\circ = Q^*$.

To complete the proof we have to show that K^* is indeed in E_K. But this follows from the fact that Q^* is a stabilizing solution. Indeed,

$$F^T + cQ^*GG = (-R^*K^{*T} + I/2) + cQ^*K^*S^*K^{*T} = -I/2.$$

An alternative proof of Theorem 5.1 characterizing the unique stationary point of $J(K)$ in E_K can be obtained using the representation of the asymptotic cost functional given in (3.9) via the filter-Riccati equation and applying the Lagrange-multipliers method. Not going into details we remark that the optimal value of the corresponding Lagrange-multipliers, denoted generically by Λ_f and the optimal value of P are given as follows:

$$\Lambda_f^* = H^T H, \qquad P^* = K^*S^*(R^*)^{-1}.$$

With these values we obtain

$$F + cP^*H^TH = -K^*R^* + I/2 + cP^*H^TH = -I/2, \tag{5.12}$$

just like in the control-Riccati case.

Remark. It can also be shown that the Hessian-matrix of $J(K)$ at $K = K^*$ is positive definit. Thus K^* is a strict local minimum of $J(.)$.

6 Minimization of $J(K)$ over E_K: a filter-Riccati approach.

Next we show, using the filter-Riccati equation representation of $J(K)$, that the unique stationary point of $J(K)$ in E_K found in the previous section in Theorem 5.1, is in fact the unique minimum point if $J(K)$ in E_K.

Theorem 6.1 *Assume that $S^* > cS^*(R^*)^{-1}H^TH(R^*)^{-1}S^*$. Then $J(K)$ has a unique minimum point in E_K given by*

$$K^* = (R^* - cS^*(R^*)^{-1}H^TH)^{-1}.$$

The corresponding cost is

$$J^* = J(K^*) = \mathrm{tr}\ (R^*S^{*-1}R^* - cH^TH)^{-1}H^TH\ .$$

Proof. First we note that if F is asymptotically stable then P is positive definite. (This well-known fact of the theory of Riccati equations is shown as follows. If $Px = 0$, then $x^TKS^*K^Tx = 0$, thus $K^Tx = 0$, consequently

$$F^Tx = \frac{x}{2}\ ,$$

contradicting the assumed stability of F.

Rewrite now the filter-Riccati equation so that R^* is replaced by $R^* - cS^*(R^*)^{-1}H^T H$ in $F = -KR^* + I/2$, a step motivated by the results of [12] or of Section 5 on the unique stationary point of $J(K)$. Set

$$F_{\mathrm{mod}} = -K\left(R^* - cS^*(R^*)^{-1}H^T H\right) + \frac{I}{2}\,.$$

Then

$$\begin{aligned} F_{\mathrm{mod}}P + PF_{\mathrm{mod}}^T + K\left(S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*\right)K^T \\ +c(KS^*(R^*)^{-1} - P)H^T H(KS^*(R^*)^{-1} - P)^T = 0\,. \end{aligned} \tag{6.1}$$

We prove that this equation in turn implies that F_{mod} is asymptotically stable. In fact, since $P > 0$ it is enough to check the controllability of the pair

$$\left(F_{\mathrm{mod}}, \left[K\left(S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*\right), (KS^*(R^*)^{-1} - P)H^T H\right]\right)\,.$$

For this we apply the celebrated P-B-H test. Consider a left eigenvector $\overline{y}^T$ of the matrix F_{mod} for which

$$\overline{y}^T\left[K\left(S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*\right), (KS^*(R^*)^{-1} - P)H^T H\right] = [0, 0]\,, \tag{6.2}$$

holds. Then, since $S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*$ is nonsingular, $\overline{y}^T K = 0$. Consequently, $\overline{y}^T PH^T H = 0$ and $\overline{y}^T F = \frac{\overline{y}^T}{2}$. Multiplying the filter-Riccati equation by $\overline{y}^T$ and $\overline{y}$ from the left and the right, respectively, we get that

$$\overline{y}^T\left(FP + PF^T + KS^*K + cPH^T HP\right)\overline{y} = 0\,,$$

and using the equalities above equation

$$\overline{y}^T P\overline{y} = 0$$

is obtained, contradicting to $P > 0$.

Next we compare the matrix P in (6.1) with $\left(R^*S^{*-1}R^* - cH^T H\right)^{-1}$. Straightforward calculation gives that – introducing the notation

$$P_{\mathrm{mod}} = P - \left(R^*S^{*-1}R^* - cH^T H\right)^{-1}$$

equation (6.1) can be rewritten as follows:

$$\begin{aligned} & F_{\mathrm{mod}}P_{\mathrm{mod}} + P_{\mathrm{mod}}F_{\mathrm{mod}}^T \\ & + \left[F_{\mathrm{mod}} + \frac{I}{2}\right]\left(R^*S^{*-1}R^* - cH^T H\right)^{-1}\left[F_{\mathrm{mod}} + \frac{I}{2}\right]^T \\ & + c(KS^*(R^*)^{-1} - P)H^T H(KS^*(R^*)^{-1} - P)^T = 0\,. \end{aligned} \tag{6.3}$$

Since the last two terms in this equation are positive semidefinit the asymptotic stability of the matrix

$$-K\left(R^* - cS^*(R^*)^{-1}H^TH\right) + \frac{I}{2}$$

implies that

$$P \geq (R^*S^{*-1}R^* - cH^TH)^{-1} \tag{6.4}$$

Since the functional $J(K)$ on E_{K} can be written in the form $J(K) = \text{ tr } PH^TH$, (6.4) implies

$$J(K) = \text{tr } PH^TH \geq \text{tr } \left(R^*S^{*-1}R^* - cH^TH\right)^{-1} H^TH.$$

Since H^TH is nonsingular, equality holds if and only if $P = \left(R^*S^{*-1}R^* - cH^TH\right)^{-1}$. Substituting this value into (6.3) we obtain that $KS^*(R^*)^{-1} = P$. Thus the unique optimal K is

$$K^* = \left(R^* - cS^*(R^*)^{-1}H^TH\right)^{-1} \ ,$$

concluding the proof of the Theorem 6.1.

Remark. Note, that in the previous proof the invertibility of the matrix H was not used. On the other hand, in order to write the LEQG cost in form of (3.9) in Proposition 3.4 the observability of the pair (H, F) is essential. Thus the above proof can also be used in the case of singular H assuming that $K \in E_{\mathrm{K}}$ and (H, F) is an observable pair to derive that

$$J(K) \geq \text{tr } (R^*S^{*-1}R^* - cH^TH)^{-1}H^TH \ .$$

At the same time, the assumption that the pair (H, F) is observable for *any* matrix $K \in E_{\mathrm{K}}$ is obviously equivalent to the invertibility of H.

In the special case of ARMA processes we have $S^* = R^*$, thus we get the following result:

Corollary 6.1 *Consider an ARMA-system and assume that $R^* > cH^TH$. Then the set E_{K} defined under (3.3) is non-empty, and $J(K)$ has a unique minimum in E_{K} given by*

$$K^* = \left(R^* - cH^TH\right)^{-1} \ ,$$

and the corresponding cost is

$$J(K^*) = \text{tr } H^TH\left(R^* - cH^TH\right)^{-1} \ .$$

The above value for the optimal K has been found for AR-processes also in [12].

To apply our general result for multi-variable linear stochastic systems consider first the special case of Theorem 6.1 when H, R^*, S^* are block-diagonal, say

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \tag{6.5}$$

and the partition of H is identical with that of R^* and S^*. Then by the above theorem, K^* is also block-diagonal with blocks K_1^* and K_2^*, say, and for $i = 1, 2$ we have

$$K_i^* = \left(R_i^* - cS_i^*(R_i^*)^{-1}H_i^T H_i\right)^{-1} . \tag{6.6}$$

If we knew beforehand that K^* is block-diagonal, a fact that apparently can not be easily shown directly, then we could restrict the minimization problem to the set of block-diagonal matrices K, say with

$$K = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix},$$

and we could decompose the original problem, and we could write

$$J(K) = J_1(K_1) + J_2(K_2).$$

The minimization of $J(K)$ over block-diagonal K-s thus can be reduced to two separate minimization problems with respect to K_1 and K_2. A direct application of the above remark gives the following result:

Corollary 6.2 *Consider the multivariable linear stochastic systems given by (2.1), and satisfying Conditions 2.1, 2.2 and 2.3. Consider the risk-sensitive identification criterion for (θ^*, Λ^*) given in (2.6) such that H is block-diagonal, see (6.5). Assume that*

$$S^* > cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*,$$

in particular

$$R_1^* > cH_1^T H_1,$$

where R^*, S^* are defined under (2.3), (2.4), and R_1^* is defined under (2.2). Then the set E_{K} defined under (3.3) is non-empty, and $J(K)$ has a unique minimum point in E_{K}, it is also block-diagonal and its $(1,1)$ *block is given by*

$$K_1^* = \left(R_1^* - cH_1^T H_1\right)^{-1} .$$

Check the following !

Remark. Note that the condition $R_1^* > cH_1^T H_1$ is scale -invariant in the sense that if the innovation is multiplied by a constant matrix, then the validity of this condition is unaffected. This follows from the fact that R_1^* is normalized by the covariance-matrix of the innovation, Λ^* , see (2.2), thus it is scale-independent. The situation is different in estimating Λ^*. In the condition $S_2^* > cS_2^*(R_2^*)^{-1}H^T H(R_2^*)^{-1}S_2^*$ with obvious definitions of R_2^* and S_2^*, see (2.3) and (2.4), we see that multiplying the innovation by a constant scalar leaves the right hand side unaffected, while the left hand side is multiplied by the square of this constant.

7 Minimization of $J(K)$ over $E_{\rm K}^{\circ}$.

So far we have considered the minimization of $J(K)$ over $E_{\rm K}$ defined in (3.3) as

$$E_{\rm K} = \{K : \ K \in D_{\rm K} \text{ and } \|c^{1/2}\mathcal{G}\|_\infty < 1\}.$$

Recall that Proposition 3.1 implies that on the set

$$\{K : \ K \in D_{\rm K} \text{ and } \|c^{1/2}\mathcal{G}\|_\infty > 1\}$$

$J(K)$ is well-defined and $J(K) = \infty$ and the definition of $J(K)$ on the set

$$E_{\rm K}^{\circ} = \{K : \ K \in D_{\rm K} \ , \|c^{1/2}\mathcal{G}\|_\infty \leq 1\} \ . \tag{7.1}$$

was extended in the following way (see (3.6)):

$$J(K) = \frac{2}{c} \liminf_{T\to\infty} \frac{1}{T} \log \mathrm{E} \left(\exp\{\frac{c}{2} \int_0^T \widetilde{x}(t)^T H^T H \widetilde{x}(t) dt\} \right) . \tag{7.2}$$

Now letting the constant c vary we will prove the following extension of Theorem 5.1:

Theorem 7.1 *Assume that $S^* - cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*$ is positive definite. Then the set $E_{\rm K}^{\circ}$ defined under (3.7) is non-empty, and $J(K)$ achieves its minimum over $E_{\rm K}^{\circ}$ at the unique minimizing K given by*

$$K^* = (R^* - cS^*(R^*)^{-1}H^T H)^{-1}.$$

The optimal cost is

$$J^* = J(K^*) = \mathrm{tr}\ (R^* S^{*-1} R^* - cH^T H)^{-1} H^T H.$$

Remark. Observe that Theorem 7.1 implies the following result which might be surprising at first sight. Consider the set

$$E_c = \left\{ c \mid \text{there exists a } K \text{ such that } \ K \in D_{\rm K} \quad \text{and} \quad \|c^{1/2}\mathcal{G}\|_\infty \leq 1 \right\} \ .$$

Then by the continuity of the H_∞-norm with respect to c and K the set E_c is an *open* interval.

Proof. Since in the course of the proof the parameter c will vary we shall express this dependence by the notations $E_{\rm K}^{\circ}(c), E_{\rm K}(c)$ etc. with their obvious meaning. In view of Theorem 6.1 it is enough to prove that if $K \in E_{\rm K}^{\circ}(c)$ but $K \notin E_{\rm K}(c)$ then $J(K,c) > J(K^*(c),c)$.

First observe that $\frac{c}{2}J(K,c)$ is a monotonically increasing function of c and if $K \in E_{\mathrm{K}}^{\circ}(c)$ then $K \in E_{\mathrm{K}}(c')$ for any $c' < c$. On the other hand Theorem 6.1 implies

$$\lim_{c' \nearrow c} J(K^*(c'), c') = J(K^*(c), c) \ , \tag{7.3}$$

as long as the inequality $S^* > cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*$ holds.

Now let $K \in E_{\mathrm{K}}^{\circ}(c)$. Then by Propositions 3.1 and 3.3

$$S^* > cS^*(R^*)^{-1}H^T H(R^*)^{-1}S^*,$$

hence (7.3) applies. Thus

$$\frac{c}{2}J(K,c) \geq \lim_{c' \nearrow c} \frac{c'}{2}J(K,c') \geq \lim_{c' \nearrow c} \frac{c'}{2}J(K^*(c'),c') = \frac{c}{2}J(K^*(c),c), \tag{7.4}$$

The first inequality follows from the monotonicity of $\frac{c}{2}J(K,c)$, the second follows since $K \in E_{\mathrm{K}}(c')$, and the last equality is just (7.3). This proves that $K^*(c)$ is optimal even in $E_{\mathrm{K}}^{\circ}(c)$.

To prove *uniqueness* we need a more delicate analysis. First note, that (7.4) implies that

$$\frac{c}{2}J(K,c) - \frac{c}{2}J(K^*(c),c) \geq \lim_{c' \nearrow c} \frac{c'}{2}\left(J(K,c') - J(K^*(c'),c')\right). \tag{7.5}$$

Take a c' with $0 < c' < c$, then by Proposition 3.4 we can write

$$\Delta J(c') := J(K,c') - J(K^*(c'),c') = \mathrm{tr}\ PH^T H - \mathrm{tr}\ P^*(c')H^T H = \mathrm{tr}\ \Delta P(c') \cdot H^T H \tag{7.6}$$

with

$$\Delta P(c') = P - P^*(c').$$

Using the inequality $\mathrm{tr}\ AB \geq \lambda_{\min}(B)\lambda_{\max}(A)$, we get

$$\Delta J(c') \geq \lambda_{\min}(H^T H)\lambda_{\max}(\Delta P(c')). \tag{7.7}$$

Thus to estimate $\Delta J(c)$ from below for $K \neq K^*(c)$, $K \in E_{\mathrm{K}}^{\circ}$ it sufficient to estimate $\Delta P(c')$, $c' < c$ from below.

For this purpose we will need the following addition to Proposition 7.1, implied by Proposition 2.3.1 of [10] or Lemma 8 of [15]:

Proposition 7.1 *Assume that $F = -KR^* + I/2$ is asymptotically stable. Then the H_∞-norm of $c^{1/2}\mathcal{G}$ is less than or equal to 1 if and only if the filter-Riccati equation:*

$$FP + PF^T + cPH^T HP + GG = 0. \tag{7.8}$$

has a real, symmetric solution P for which the matrix $F + cPH^T H$ is stable. Moreover, this solution is unique and positive definite.

The following simple identity for the comparison of two solutions of the filter-Riccati equation corresponding to different K matrices is a direct consequence of the Proposition above. Assume that $K_1, K_2 \in E_{\mathrm{K}}^{\circ}(c')$ and denote by $P_1 = P_1(c'), P_2 = P_2(c')$ the corresponding unique, symmetric, positive definite solutions of the filter-Riccati equation (7.8). Then, using the notation $F_1 = -K_1 R^* + \frac{I}{2}$, and $F_2 = -K_2 R^* + \frac{I}{2}$, we get by direct computation, dropping the argument c' for a moment:

$$\begin{aligned} &\left(F_1 + c' P_2 H^T H\right)(P_1 - P_2) + (P_1 - P_2)\left(F_1 + c' P_2 H^T H\right)^T + \\ &\quad + c'(P_1 - P_2) H^T H (P_1 - P_2) + K_1 S^* K_1^T - K_2 S^* K_2^T \\ &\qquad - (K_1 - K_2) R^* P_2 - P_2 R^* (K_1 - K_2)^T = 0\,. \end{aligned} \tag{7.9}$$

Apply this identity choosing $K_1 = K \in E_{\mathrm{K}}^{\circ}(c')$, $K_2 = K^*(c') \in E_{\mathrm{K}}(c')$. (Note that the condition on K_2 has been deliberately strengthened). Then

$$P_1 = P_1(c') = P, \qquad P_2 = P_2(c') = P^*(c') = K^*(c') S^* (R^*)^{-1}, \qquad F_1 = F = F(K).$$

Thus the following Riccati equation is obtained for $P - P^*(c') = \Delta P(c')$:

$$\begin{aligned} &\left(F + c' P^*(c') H^T H\right) \Delta P(c') + \Delta P(c') \left(F + c' P^*(c') H^T H\right)^T + \\ &+ c' \Delta P(c') H^T H \Delta P(c') + \left(K - K^*(c')\right) S^* \left(K - K^*(c')\right)^T = 0. \end{aligned} \tag{7.10}$$

We show that (7.10) implies that $F + c' P^*(c') H^T H$ is asymptotically stable for any $K \in E_{\mathrm{K}}^{\circ}(c')$. Indeed, first note that

$$F + c' P^*(c') H^T H = -\left(K - K^*(c')\right) R^* - \frac{I}{2} \tag{7.11}$$

in view of (5.12). Now, if ξ is a right eigenvector of $\left(F + c' P^*(c') H^T H\right)^T$ with eigenvalue λ, then (7.10) implies that

$$2 \ \mathrm{Re}\ \lambda\ \bar{\xi}^T \Delta P(c') \xi + \bar{\xi}^T \left(K - K^*(c')\right) R^* \left(K - K^*(c')\right)^T \xi \le 0\,.$$

Now if we had $\mathrm{Re}\ \lambda \ge 0$ then $(K - K^*(c'))^T \xi = 0$ should hold. Multiplying (7.11) by $\bar{\xi}^T$ from the left we obtain that $\lambda = -\frac{1}{2}$, which is a contradiction. Thus $\mathrm{Re}\ \lambda < 0$.

Let denote by $\overline{\Delta P(c')}$ the solution of the Lyapunov-equation obtained from (7.10) by removing the quadratic term:

$$\begin{aligned} &\left(F + c' P^*(c') H^T H\right) \overline{\Delta P(c')} + \overline{\Delta P(c')} \left(F + c' P^*(c') H^T H\right)^T \\ &\qquad + \left(K - K^*(c')\right) S^* \left(K - K^*(c')\right)^T = 0. \end{aligned} \tag{7.12}$$

Then obviously,

$$\Delta P(c') \geq \overline{\Delta P(c')} \,. \tag{7.13}$$

Now we are going to take the limit for $c' \nearrow c$. Then

$$\overline{\Delta P(c')} \to \overline{\Delta P(c)} \,.$$

Obviously $\overline{\Delta P(c)} \neq 0$ for $K \neq K^*(c)$, and $\overline{\Delta P(c)}$ is positive semidefinit for $K \in E^\circ_{\mathrm{K}}(c)$ due to the asymptotic stability of $F + cP^*(c)H^T H$, and thus $\lambda_{\max}(\Delta P(c)) > 0$. We get that $\Delta J(c) > 0$ for $K \neq K^*(c)$, as stated. Thus Theorem 7.1 is proved.

Corollary 7.1 *Consider an ARMA-system and assume that $R^* > cH^T H$. Then the set E°_{K} defined under (3.7) is non-empty, and $J(K)$ achieves its minimum over E°_{K} at the unique minimizing K given by*

$$K^* = \left(R^* - cH^T H\right)^{-1} \,,$$

and the corresponding cost is

$$J(K^*) = \mathrm{tr}\ \left(R^* - cH^T H\right)^{-1} H^T H \,.$$

Similarly the minimization of the risk-sensitive identification criterion for multi-variable linear stochastic systems given in Corollary 6.2 can be extended to cover minimization over E°_{K}:

Corollary 7.2 *Consider the multi-variable linear stochastic systems given by (2.1), and satisfying Conditions 2.1, 2.2 and 2.3. Consider the risk-sensitive identification criterion for (θ^*, Λ^*) given in (2.6) such that H is block-diagonal, see (6.5). Assume that*

$$S^* > cS^*(R^*)^{-1} H^T H (R^*)^{-1} S^*,$$

in particular

$$R_1^* > cH_1^T H_1,$$

where R^*, S^* are defined under (2.3), (2.4), and R_1^* is defined under (2.2). Then the set the set E°_{K} defined under (3.7) is non-empty, and $J(K)$ achieves its minimum over E°_{K} at a unique minimizing K^*, it is also block-diagonal and its $(1,1)$ block is given by

$$K_1^* = \left(R_1^* - cH_1^T H_1\right)^{-1} \,.$$

Acknowledgment

The first author expresses his thanks to Jan van Schuppen for introducing him to LEQG-control and valuable discussions.

References

[1] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive algorithms and stochastic approximations.* Springer-Verlag, Berlin, 1990.

[2] A. Bensoussan and J.H. van Schuppen. Optimal control of partially observable stochastic systems with an exponential-of-integral performance index. *SIAM J. Control Optim.*, 23:599–613, 1985.

[3] L. Gerencsér, Rate of convergence of recursive estimators. *SIAM J. Control and Optimization*, 30(5):1200–1227, 1992.

[4] L. Gerencsér, A representation theorem for the error of recursive estimators. In *Proc. of the 31st IEEE Conference on Decision and Control, Tucson*, pages 2251–2256, 1992.

[5] D. H. Jacobson, Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games, *IEEE Tr. Automatic Control*, 18:124-131, 1973

[6] , P. R. Kumar, *Stochastic optimal control and differential games*, Dept. of System Science and Mathematics, Washington University, St. Louis, MO, USA, 1977

[7] P. R. Kumar and J. H. van Schuppen, On the optimal control of stochastic systems with an exponential-of-integral performance index, *J. Math. Anal Appl.*, 80:312-332, 1981.)

[8] L. Ljung and T. Söderström. *Theory and practice of recursive identification.* The MIT Press, 1983.

[9] H. Lütkepohl, *Handbook of matrices*, Wiley and Sons, 1996.

[10] D. Mustafa and K. Glover, *Minimum Entropy H_∞ Control.* Lecture Notes in Control and Information Sciences, Vol. 146. Springer Verlag, 1990.

[11] T. Söderström and P. Stoica. *System identification.* Prentice Hall, 1989.

[12] A. Stoorvogel and J. H. van Schuppen. *System identification with information theoretic criteria.* in *From Identification to Learning: the science of learning models from data*, Eds. S. Bittanti and G. Picci, NATO ASI Series F: Computer and Systems Sciences, Vol. 153, 1996, pp. 289-338.

[13] 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 Tr. Automatic Control*, 19:358-366, 1974.

[14] P. Whitle, Risk-sensitive Linear/Quadratic/Gaussian control, *Adv. Appl. Prob.*, 13:764-777, 1981.

[15] J. C. Willems, Least squares stationary optimal control and the algebraic Riccati equation, *IEEE Transactions of Automat. Contr.* 18(6):621–634, 1971.