Centrum voor Wiskunde en Informatica

REPORT *RAPPORT*

MAS

Modelling, Analysis and Simulation

Modelling, Analysis and Simulation

An algebraic method for system reduction of stationary Gaussian systems

D. Jibetean, J.H. van Schuppen

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

AN ALGEBRAIC METHOD FOR SYSTEM REDUCTION
OF STATIONARY GAUSSIAN SYSTEMS

Dorina Jibetean and Jan H. van Schuppen

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract: System identification for a particular approach reduces to system reduction, determining for a system with a high state-space dimension a system of low state-space dimension. For Gaussian systems the problem of system reduction is considered with the divergence rate criterion. The divergence or Kullback-Leibler pseudo-distance corresponds to the expected value of the negative natural logarithm of the likelihood function. System reduction for Gaussian systems is thus a certainty equivalent way of maximum likelihood identification. An algebraic method is proposed for system reduction. The results are a theorem that this problem reduces to an infimization problem for a rational function for which programs are available and a procedure for computing the best approximant w.r.t. the divergence rate criterion. As illustration two examples of system reduction are presented.

1. INTRODUCTION

The aim of this paper is to introduce an algebraic approach to system reduction for Gaussian systems by the divergence rate criterion.

The motivation for this paper is system identification of Gaussian systems. A finite-dimensional Gaussian system is a linear system driven by a Gaussian white noise process. In this paper attention is limited to discrete-time systems. As is well known, a stationary Gaussian system is a mathematical model for an observed stationary Gaussian process. The system identification problem is to construct from observed data and from assumptions a mathematical model, here a Gaussian system, such that the observed processes of the system approximate the observed data as well as possible according to an approximation criterion.

Methods of system identification for Gaussian systems often used include the maximization of the likelihood function, the subspace identification algorithm, and the least-squares method. The divergence between two probability measures is a well known pseudo-distance. It equals the expectation of the negative of the natural logarithm of the likelihood function. The divergence rate is derived from divergence and is needed because of the consideration of a stationary process.

The approximation problem of system identification is one of the major problems of this area. The main questions of parameter estimation include: How to find the global infimum? How to derive the first-order conditions? How to compute the local minima? How many local minima are there? Is the global minimum unique?

The aim of the paper is: (1) To present an algebraic approach and an algorithm for the infimization of the divergence rate criterion of Gaussian systems. (2) To show for several low order Gaussian systems that system reduction leads for the divergence rate criterion to two or more local minima. These examples and the method have serious implications for system identification of Gaussian systems by the maximum likelihood method. Though it is known from theoretical investigations and from numerical experiments with examples of system identification problems that two or more local minima exist, the consequences of this for system identification practice seem not to be widely known.

The results of the paper include a procedure to determine the best approximant w.r.t. the divergence rate criterion by an algebraic method. Determining the best approximant is proven to be equivalent to infimization of a rational function for which recently an algorithm was determined by the first-named author, see (Jibetean, 2001; Jibetean, 2003). The approach is illustrated in Example 2 with the reduction from a third order Gaussian system to a second order one. The set of local minima is not completely determined in this case although an upper bound on its cardinality is provided. Example 1 treats model reduction for a Gaussian system of state-space dimension 2 to one of state-space dimension 1. In this case there are two potential minima, one is the global minimum and the other a local one. The criterion values are quite close.

The novelty of the paper is in: (1) The combination of the divergence rate criterion for system reduction with algebraic methods; (2) The equivalence of the system reduction problem w.r.t. the divergence rate criterion to optimization of a rational function of the system parameters; (3) The application of algorithms for infimization of rational functions (based on LMI relaxations) to system reduction w.r.t. the divergence rate criterion; (4) Two examples which illustrate that two or more local minima of the criterion exist.

In comparison with the literature, algebraic approaches for computing *globally* best approximants have been applied to system reduction w.r.t. the H_2 criterion, see e.g. (Hanzon and Maciejowski, 1996), (Jibetean and Hanzon, 2002). The main interest to system identification of the proposed method is that it provides an algebraic way to determine the global minimum of the criterion and to circumvent the case of multiple local minima.

A description of the contents of the paper follows. The next section contains a short problem formulation. This section is best read in combination with the appendices. Section 3 presents the procedure for system reduction via divergence rate infimization. The

algebraic method is presented in Section 4. Examples are provided in Section 5. Conclusions are stated in the last section. Appendix A contains notation and terminology on linear systems, Appendix B on Gaussian systems, and Appendix C the formulas for the divergence rate of stationary Gaussian systems.

This paper is a sequel to those of (Stoorvogel and van Schuppen, 1996; Stoorvogel and van Schuppen, 1998).

2. PROBLEM FORMULATION

The motivating engineering problem is to determine a simple mathematical model for a time series. One speaks of the system identification problem or of the approximate realization problem. Examples of such a problem are the modeling of a signal in a noisy communication channel, of messages in a digital communication network, and of the traffic flow on a motorway.

Mathematical notation for the problem is summarized below. See the appendices for further details. Let (Ω, F) be a measurable space and $T = \mathbb{Z}$ denote the time index set and let $\mathbb{N} = \{0, 1, \ldots\}$ denote the set of natural numbers. Let P_1 be a probability measure on (Ω, F) induced by a stationary Gaussian process $y : \Omega \times T \to \mathbb{R}^p$ with zero mean value function and covariance function $W : T \to \mathbb{R}^{p \times p}$.

A time-invariant finite-dimensional Gaussian system on a probability space (Ω, F, P) is a stochastic system with representation

$$
\begin{aligned}
x(t+1) &= Ax(t) + Bv(t), \\
y(t) &= Cx(t) + Dv(t), \\
&(A, B, C, D) \in LSP(p, n, p), \\
&x : \Omega \times T \to \mathbb{R}^n, \ y : \Omega \times T \to \mathbb{R}^p,
\end{aligned}
$$

see Appendix A for the full specification of the system. If the parameters of the system are in the set $SGSP_{min}(p, n, p)$ then the output process is a stationary Gaussian process. The probability measure induced by this system on the output process y is denoted by $P(q)$ where $q \in QD$ represents the parameter of a selected parametrization.

In this paper attention is restricted to the approximation problem of the system identification procedure.

Procedure 1. (1) Determine from a finite time series a high-order Gaussian system.
(2) System reduction: Determine from a high-order Gaussian system a low-order Gaussian system.

In this paper attention for the approximation problem is restricted to the divergence rate criterion. The concept of divergence of two probability measures is used in information theory. In probability theory divergence corresponds to the Kullback-Leibler measure, see (Cover and Thomas, 1991). For a stationary stochastic process the concept of divergence rate of two probability measures has been defined. In Section 3 an expression is provided for the divergence rate of two measures induced by stationary Gaussian processes which are outputs of two time-invariant finite-dimensional Gaussian systems. Denote this divergence rate by $D_r(P_1 \| P_2)$.

Let $n_2 \in \mathbb{N}$ denote an upper bound on the dimension of the Gaussian system to be determined.

Problem 1. Solve

$$\inf_{n \leq n_2, \, q \in SGSP_{min}(p,n,p)} D_r(P_1 \| P(q)). \quad (1)$$

The problem involves establishing whether or not a minimum exists, if a minimum exists to characterize the set of minima, and to construct a procedure to compute a minimum or to approximate an infimum.

3. PROCEDURE FOR INFIMIZATION OF DIVERGENCE RATE

Recall the formula for the divergence rate of two measures induced by stationary Gaussian processes which are outputs of two time-invariant finite-dimensional Gaussian systems

System 1 $n_1 \in \mathbb{N}$,
$(A_1, B_1, C_1, D_1) \in SGSP_{min}(p, n_1, p)$,
System 2 $n_2 \in \mathbb{N}$,
$(A_2, B_2, C_2, D_2) \in SGSP_{min}(p, n_2, p)$.

The expression is available from the literature (Stoorvogel and van Schuppen, 1996; Stoorvogel and van Schuppen, 1998) in terms of a realization, (A_4, B_4, C_4, D_4), of the series interconnection of System 3 and System 1, where System 3 is the inverse of System 2. The relation between System 2 and System 3 is expressed by $n_3 = n_2$ and

$$(A_3, B_3, C_3, D_3)$$
$$= (A_2 - B_2 D_2^{-1} C_2, B_2 D_2^{-1}, -D_2^{-1} C_2, D_2^{-1})$$
$$\in SGSP_{min}(p, n_3, p).$$

Procedure 2. The divergence rate associated with the Systems 1 and 2 is computed by the following steps:

(1) Construct (A_4, B_4, C_4, D_4) according to the formulas $n_4 = n_1 + n_3$,

$$(A_4, B_4, C_4, D_4)$$
$$= \left(\begin{pmatrix} A_1 & 0 \\ B_3 C_1 & A_3 \end{pmatrix}, \begin{pmatrix} B_1 \\ B_3 D_1 \end{pmatrix}, \right.$$
$$\left. \begin{pmatrix} D_3 C_1 & C_3 \end{pmatrix}, D_3 D_1 \right) \in SLSP(p, n_4, p).$$

(2) Solve the discrete-time Lyapunov equation for the matrix $Q_4 \in \mathbb{R}^{n_4 \times n_4}$,

$$Q_4 = A_4 Q_4 A_4^T + B_4 B_4^T. \quad (2)$$

(3) Calculate

$$f_c = D_r(P_1 \| P_2) \quad (3)$$
$$= \frac{1}{2} \text{tr}(C_4 Q_4 C_4^T + D_4 D_4^T - I)$$
$$- \frac{1}{2} \ln \det(D_4 D_4^T).$$

Algorithm 1. Infimization of the divergence rate of stationary Gaussian processes.
Input: System 1 representing the first probability measure and $n_2 \in \mathbb{N}$, the desired order of the approximant.
Output: System 2 representing the probability measure associated to the approximant

(1) Parametrize System 3 by a canonical parametrization map $f_p : QD \rightarrow SGSP_{min}(p, n_3, p)$,

$$q \overset{f_p}{\mapsto} (A_3(q), B_3(q), C_3(q), D_3(q)).$$

Note that $QD \subseteq \mathbb{R}^r$, where r is the dimension of the manifold $SGSP_{min}(p, n, p)$. Here f_q is restricted to functions which are rational in the entries of the parameter vector q.

(2) Determine, if it exists, a parameter value $\hat{q}_3 \in QD$ such that

$$\hat{q}_3 = \text{argmin}_{q \in QD} f_c(q), \quad (4)$$
$$f_c(q) := D_r(P_1 \| P_2(q)), \quad (5)$$

where $f_c(q)$ is determined according to Procedure 2.

(3) Set $(\hat{A}_3, \hat{B}_3, \hat{C}_3, \hat{D}_3) = f_p(\hat{q}_3)$ according to the parameterization map f_p.

(4) Compute the approximant System 2 according to

$$(\hat{A}_2, \hat{B}_2, \hat{C}_2, \hat{D}_2) = \left(\hat{A}_3 - \hat{B}_3 \hat{D}_3^{-1} \hat{C}_3, \quad (6) \right.$$
$$\left. \hat{B}_3 \hat{D}_3^{-1}, -\hat{D}_3^{-1} \hat{C}_3, \hat{D}_3^{-1} \right).$$

4. ALGEBRAIC METHOD

For the divergence infimization an algebraic method will be used. The *algebraic method* refers to the use of abstract algebra, computer algebra, and the use of the

computer programs like MAPLE and MATHEMAT-ICA. The difficulties to be overcome in the algebraic methods are to organize the calculations and to find an approach that is of low complexity.

Theorem 1. Consider the infimization problem

$$\inf_{q \in QD} f_c(q),$$

where $f_c(q)$ is computed by Procedure 2 and the matrices (A_3, B_3, C_3, D_3) depend on the parameter vector $q \in QD$.

(a) The infimization of the criterion with respect to the matrix C_3 is reached at the matrix

$$C_3 = -D_3 C_1 Q_2 Q_3^{-1}, \text{ where,} \qquad (7)$$
$$Q_4 = \begin{pmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{pmatrix} \in \mathbb{R}^{n_4 \times n_4}$$
is the solution of (2),
$$Q_2 \in \mathbb{R}^{n_1 \times n_2}, \ Q_3 \in \mathbb{R}^{n_2 \times n_2}.$$

Hence the criterion depends on the matrices A_3, B_3 and D_3.

(b) The infimization with respect to $D_3 \in \mathbb{R}^{p \times p}$ is reached for D_3 satisfying

$$D_3^T D_3 = M^{-1}, \qquad (8)$$

and the criterion simplifies to,

$$f_c(q) = -\frac{1}{2} \ln \det(D_1^T M^{-1} D_1)$$
$$= \frac{1}{2} \ln \det M - \ln \det D_1 \ , \text{ where}$$

$$M = C_1 \left(Q_1 - Q_2 Q_3^{-1} Q_2^T \right) C_1^T + D_1 D_1^T.$$

The simplified criterion is a natural logarithm of a function which is a rational function with respect to entries of the (A_3, B_3). Thus the infimization problem is reduced to an infimization problem for a rational function.

PROOF. See (Jibetean, 2003). □

Procedure 3. (1) Select a parameterization for the matrices of System 3, A_3, B_3, C_3, and D_3, see Algorithm 1. In view of Theorem 1, choose a parameterization with C_3, D_3 fully parametrized, independent from A_3, B_3. For example, the control canonical form gives such a parameterization.

(2) Solve by computer algebra the discrete-time Lyapunov equation (2) for the symbolic matrix $Q_4(q)$.

(3) Calculate the value of the criterion according to formula (3). The criterion f_c is the sum of a rational function and of a natural logarithm of

the parameters of the system matrices A_3, B_3, C_3, and D_3.

(4) Apply the reduction technique formulated in Theorem 1 to solve analytically for the matrices C_3 and D_3 and to derive the simplified formula for the criterion. There remains then an infimization problem for a rational function, $\det M$.

(5) Determine the value of the infimum. If, moreover, the infimum is attained, i.e. the global minimum exists, then determine its location as well. For this use the approach of (Jibetean, 2001; Jibetean, 2003). Once M is determined, compute D_3 by (8) and C_3 by (7).

(6) If this is of interest then the infimization of the rational function also can provide information on the local minima. Derive the first order conditions of the simplified criterion with respect to the elements of the parameter vector $q \in QD$. Computer algebra provides programs for this.

(7) Determine all solutions in the set of the real numbers of the equation obtained by setting to zero the first derivative of the criterion with respect to the parameter vector. This is the most difficult and demanding part of the procedure.

(8) Calculate for each solution the second derivative of the criterion. Discard all points for which the second derivative is not positive semi-definite.

(9) For each of the remaining points calculate the value of the criterion $f_c(q)$. By comparing the different values numerically determine the global minimum or the set of global minima if there exist two or more parameter vectors which attain exactly the same value.

Note that steps (6)-(9) are optional. They should be executed only if there is interest in local minima.

The problem of finding the global optimum of a rational function at step (5), is solved in (Jibetean, 2001; Jibetean, 2003) by constructing an LMI relaxation of the original problem. The LMI relaxation returns in general a lower bound on the sought infimum but it is sharp and the sharpness of the lower bound can be checked under certain conditions. In case it is not sharp one constructs a sequence of LMI relaxations of the original problem which returns a sequence of increasing lower bounds, converging to the infimum of the rational function. Then the computations are more expensive.

5. EXAMPLES

Example 1. Consider a Gaussian system of order 2 with representation in the control canonical form as

$$A_1 = \begin{pmatrix} -0.4 & -0.32 \\ 1 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$
$$C_1 = \begin{pmatrix} 0 & -0.28 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1 \end{pmatrix}.$$

An approximant will be determined in the form of a Gaussian system of order 1, according to the divergence rate criterion. The class of Gaussian systems in which an approximant is to be sought is taken to be $SGSP_{min}(1,1,1)$. This class is parameterized by the control canonical form, hence

$$(A_2, B_2, C_2, D_2) = (a_2, 1, c_2, d_2).$$

If $d_2 > 0$, $|a_2| < 1$, $|a_2 - c_2 d_2^{-1}| < 1$, $c_2 \neq 0$, then $(A_2, B_2, C_2, D_2) \in SGSP_{min}(1,1,1)$.

Construct the quadruple, in control canonical form

$$(A_3, B_3, C_3, D_3) = (a_2 - c_2 d_2^{-1}, 1, -c_2 d_2^{-2}, d_2^{-1})$$

and compute the criterion to be minimized. As remarked, the optimum with respect to c_3 and d_3 can be computed analytically. The criterion becomes

$$f_c(q) = -\frac{1}{2} \ln \left(\frac{-34 \left(25 + 10\, a_3 + 8\, a_3{}^2 \right) \dots}{\left(731\, a_3{}^2 + 1801\, a_3 + 19500 \right) \dots} \right)$$

The critical points equation with respect to a_3 is a univariate polynomial in a_3 whose roots are computed by numerical approximation. It turns out that in the stability region there exist two points of minimum of the criterion $f_c((\hat{a}_3, \hat{b}_3, \hat{c}_3, \hat{d}_3))$ such that

$$f_c((0.6353, 1, 0.1059, 0.9631)) = 0.0376,$$
$$f_c((-0.7835, 1, -0.1269, 0.9693)) = 0.0312.$$

In consequence, the second point is a global minimum, while the first returns a local minimum, although their values are close. As in Step (4) of Algorithm 1, compute the approximants

$$(\hat{a}_2, \hat{b}_2, \hat{c}_2, \hat{d}_2) = (0.5253, 1.0383, -0.1142, 1.0383),$$
$$\text{respectively } (-0.6525, 1.0317, 0.1351, 1.0317).$$

However the two approximant systems have a very different behavior. Below, the impulse response of the global approximant (green), respectively the local approximant (red), together with the impulse response of the original system (blue) are plotted.

Example 2. Consider also a model reduction from order 3 to order 2, for the system

$$A_1 = \begin{pmatrix} -1/4 & 1/2 & 1/3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$
$$C_1 = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 2 \end{pmatrix}.$$
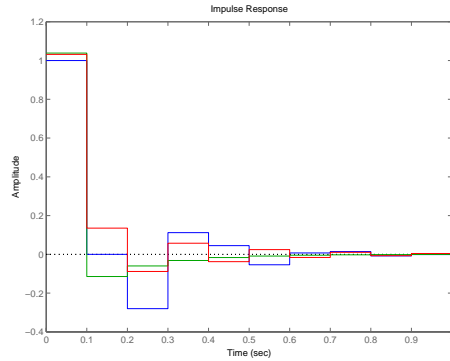
The approximant is taken in the control canonical form, parameterized by $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \delta$. After optimizing analytically with respect to $\gamma_1, \gamma_2, \delta$, an optimization of a logarithm of a rational function remains

$$-\frac{1}{2} \ln \left(\frac{\left(5640\alpha_1{}^3 + 85896\alpha_1{}^2 + 201240\alpha_1 + \dots \right)}{\left(376\alpha_2{}^3 - 618\alpha_2{}^2 + \dots \right)} \right),$$

which reduces, due to the monotonicity of the logarithm function, to optimization of a rational function. Using Gröbner bases methods for the first order conditions, one can notice that the function above has at most 100 complex critical points, including multiplicities. The computer failed to compute all of them. However, the methods of (Jibetean, 2001; Jibetean, 2003) can be employed for computing global optima of rational functions.

6. CONCLUSIONS

The main result of the paper is Procedure 3 with an algebraic method for infimization of the divergence rate between a Gaussian system and a class of such systems of lower state-space dimension. Theorem 1 establishes that the infimization problem reduces to an infimization problem for a rational function. Two examples illustrate the approach. In general a system reduction problem with this criterion and, by analogy, the parameter estimation with the likelihood function, will have many local minima. Further research is required to make the algebraic method more efficient and to streamline the computer algebra.

The authors advise the use of the proposed approach but caution the reader that at the current state of computer algebra, the method can handle effectively only low order systems. The authors also recommend for system identification of multi-output Gaussian systems the subspace identification algorithm which is based on stochastic realization theory.

5

ACKNOWLEDGMENTS

Appendix A. LINEAR SYSTEMS

In the body of the paper concepts and results for time-invariant finite-dimensional linear systems are needed.

A *discrete-time time-invariant finite-dimensional linear system* is a dynamical system with the representation

$$x(t+1) = Ax(t) + Bu(t), \quad x(t_0) = x_0,$$
$$y(t) = Cx(t) + Du(t),$$

where $T = \{t_0, t_0 + 1, \ldots\}$ is called the time axis, $n, m, p \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$ is called the initial state, $u : T \to \mathbb{R}^m$ is called the input function, $x : T \to \mathbb{R}^n$ is called the state function, $y : T \to \mathbb{R}^p$ is called the output function, and $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times m}$. The parameters of this system will be denoted by

$$(A, B, C, D) \in LSP(p, n, m).$$

Denote the reachability matrix and the observability matrix of this system respectively by

$$\mathcal{R}(A, B) = \begin{pmatrix} B & AB & \ldots & A^{n-1}B \end{pmatrix} \in \mathbb{R}^{n \times mn},$$

$$\mathcal{O}(A, C) = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} \in \mathbb{R}^{np \times n}.$$

It is said that (A, B) is a *reachable pair* if $\mathrm{rank}(\mathcal{R}(A, B)) = n$ and that (A, C) is an *observable pair* if $\mathrm{rank}(\mathcal{O}(A, C)) = n$. Denote the spectrum of the matrix $A \in \mathbb{R}^{n \times n}$ by $\mathrm{spec}(A)$ and let $\mathbb{C}^- = \{\lambda \in \mathbb{C} | \, |\lambda| < 1\}$ denote the interior of the unit disc in the complex plane. Define the subclasses of linear systems

$$LSP_{min}(p, n, m)$$
$$= \left\{ \begin{array}{l} (A, B, C, D) \in LSP(p, n, m)| \\ (A, B) \text{ reachable pair,} \\ (A, C) \text{ observable pair} \end{array} \right\},$$

$$SLSP(p, n, p)$$
$$= \left\{ \begin{array}{l} (A, B, C, D) \in LSP(p, n, p)| \\ \mathrm{rank}(D) = p, \, \mathrm{spec}(A) \subset \mathbb{C}^-, \\ \mathrm{spec}(A - BD^{-1}C) \subset \mathbb{C}^- \end{array} \right\},$$
$$SLSP_{min}$$
$$= SLSP(p, n, p) \cap LSP_{min}(p, n, m).$$

Appendix B. GAUSSIAN SYSTEMS

A time-invariant finite-dimensional *Gaussian system* (without inputs) is a stochastic system with representation

$$x(t+1) = Ax(t) + Bv(t), \qquad \text{(B.1)}$$
$$y(t) = Cx(t) + Dv(t), \qquad \text{(B.2)}$$

where $r, n, p \in \mathbb{N}$, $p \geq 1$, $v : \Omega \times T \to \mathbb{R}^r$ is a Gaussian white noise process, thus an independent sequence of random variables with for each $t \in T$, $v(t) \in G(0, V)$ ($v(t)$ has a Gaussian probability distribution function with parameters 0 and V), $V \in \mathbb{R}^{r \times r}$, $V = V^T > 0$; $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$, $C \in \mathbb{R}^{p \times n}$, $D \in \mathbb{R}^{p \times r}$; $x : \Omega \times T \to \mathbb{R}^n$, $y : \Omega \times T \to \mathbb{R}^p$ are stochastic processes satisfying the recursions (B.1,B.2).

Below a canonical form is used for Gaussian systems with respect to the covariance function of the output of the Gaussian system. For this purpose the reader is reminded of the theorem that a Gaussian system is a minimal stochastic realization of its output process iff it is stochastically observable and stochastically reconstructible, see (Lindquist and Picci, 1996). Consider a Gaussian system that is stable, with $\mathrm{spec}(A) \subset \mathbb{C}^-$. Let $Q \in \mathbb{R}^{n \times n}$ be the solution of the discrete Lyapunov equation $Q = AQA^T + BB^T$, and let $G = AQC^T + BVD^T \in \mathbb{R}^{n \times p}$. Then the Gaussian system is a minimal stochastic realization of its output process iff (A, C) is an observable pair and (A, G) is an observable pair. A time-invariant finite-dimensional Gaussian system is said to be a *Kalman realization* if in addition to being of minimal state-space dimension it satisfies $r = p$, $\mathrm{rank}(D) = p$, $\mathrm{spec}(A) \subset \mathbb{C}^-$, and $\mathrm{spec}(A - BD^{-1}C) \subset \mathbb{C}^-$.

Define the set of parameters of Gaussian systems with $p, n, r, \in \mathbb{N}$ by

$$SGSP_{min}(p, n, p)$$
$$= \left\{ \begin{array}{l} (A, B, C, D) \in SLSP(p, n, p) \, | \\ V = I, \, (A, B) \text{ reachable pair,} \\ (A, C), \, (A, G) \text{ observable pairs,} \end{array} \right\}.$$

Appendix C. DIVERGENCE RATE

The *divergence* or the *Kullback-Leibler pseudo-distance* on the set of probability measures of a measurable space (Ω, F) is defined by the formula

$$D(P_1 \| P_2) = E_Q[r_1 \ln(\frac{r_1}{r_2}) I_{(r_2 > 0)}]$$
$$= \int_\Omega r_1(\omega) \ln\left(\frac{r_1(\omega)}{r_2(\omega)}\right) I_{(r_2(\omega) > 0)} Q(d\omega),$$

where Q is a σ-finite measure on (Ω, F) such that

$$P_1 \ll Q, \; \frac{dP_1}{dQ} = r_1, \;\; P_2 \ll Q, \; \frac{dP_2}{dQ} = r_2,$$

see (Cover and Thomas, 1991, Ch. 16) and (Stoorvogel and van Schuppen, 1996, Def. C.7).

Let $y_1 : \Omega \times T \to \mathbb{R}^p$ be a stationary stochastic process on $T = \mathbb{Z}$. Denote by P_1, P_2 two measures for process y_1 on $(\mathbb{R}^p)^T$. The *divergence rate* between P_1, P_2 is defined by the formula

$$D_r(P_1 \| P_2) = \lim_{n \to \infty} \frac{1}{2n+1} D(P_1|_{[-n,n]} \| P_2|_{[-n,n]}),$$
(C.1)

if the limit exists, where $P_1|_{[-n,n]}, P_2|_{[-n,n]}$ denote the restrictions of P_1, P_2 respectively to probability measures of processes defined on the time index set $\{-n, \ldots, -1, 0, 1, \ldots, n\}$, see (Stoorvogel and van Schuppen, 1996, Def. E.4). It is shown in (Stoorvogel and van Schuppen, 1998), based on a theorem of (Stoorvogel and van Schuppen, 1996), that $D_r(P_1 \| P_2)$, the divergence rate of two probability measures induced by the output processes of two time-invariant finite-dimensional Gaussian systems, can be expressed in terms of a realization of the series interconnection between the first system and the inverse of the second system, as described in Procedure 2.

Appendix D. REFERENCES

Cover, T.M. and J.A. Thomas (1991). *Elements of information theory*. John Wiley & Sons. New York.

Hanzon, B. and J.M. Maciejowski (1996). Constructive algebra methods for the L_2−problem for stable linear systems. *Automatica* **32**(12), 1645–1657.

Jibetean, D. (2001). Global optimization of rational multivariate functions. Report PNA-R0120. CWI. Amsterdam.

Jibetean, D. (2003). Algebraic optimization with applications in system theory. PhD thesis. Vrije Universiteit Amsterdam.

Jibetean, D. and B. Hanzon (2002). Linear matrix inequalities for global optimization of rational functions and H_2 optimal model reduction. In: *Proc. of the 15th International Symposium on MTNS* (D.S. Gilliam and J. Rosenthal, Eds.).

Lindquist, A. and G. Picci (1996). Geometric methods for state space identification. In: *Identification, adaption, learning* (S. Bittanti and G. Picci, Eds.). pp. 1–69. Springer. London.

Stoorvogel, A.A. and J.H. van Schuppen (1996). System identification with information theoretic criteria. In: *Identification, adaptation, learning* (S. Bittanti and G. Picci, Eds.). pp. 289–338. Springer. Berlin.

Stoorvogel, A.A. and J.H. van Schuppen (1998). Divergence rate approximation of a stationary gaussian process by the output of a gaussian system. In: *Mathematical Theory of Networks and Systems* (A. Beghi, L. Finesso and G. Picci, Eds.). Il Poligrafo. Padova. pp. 879–882.