



Centrum voor Wiskunde en Informatica

REPORTRAPPORT

Approximation problems with the divergence criterion for Gaussian variables and Gaussian processes

A.A. Stoorvogel and J.H. van Schuppen

Department of Operations Research, Statistics, and System Theory

BS-R9616 1996

Report BS-R9616
ISSN 0924-0659

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

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications.

SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

Copyright © Stichting Mathematisch Centrum
P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Approximation Problems with the Divergence Criterion for Gaussian Variables and Gaussian Processes

A.A. Stoorvogel*

*Department of Mathematics and Computing Science, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
wscواس@win.tue.nl*

J.H. van Schuppen

*CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
J.H.van.Schuppen@cwi.nl*

Abstract

System identification for stationary Gaussian processes includes an approximation problem. Currently the subspace algorithm for this problem enjoys much attention. This algorithm is based on a transformation of a finite time series to canonical variable form followed by a truncation. There is no proof that this algorithm is the optimal solution to an approximation problem with a specific criterion. In this paper it is shown that the optimal solution to an approximation problem for Gaussian random variables with the divergence criterion is identical to the main step of the subspace algorithm. An approximation problem for stationary Gaussian processes with the divergence criterion is formulated.

AMS Subject Classification (1991): 93E12, 62H20, 60G15, 94A17.

Keywords and Phrases: System identification, Gaussian processes, divergence, canonical variable form, subspace identification algorithm.

Note: Report has been submitted for publication elsewhere.

1 Introduction

The motivation for this paper is the approximation problem of system identification. System identification addresses the problem of determining a mathematical model in the form of a dynamic system or a control system of which the input and output signal are an approximation of a time series. Of the several subproblems of system identification one is the actual approximation problem. In this paper attention is restricted to the approximation of a time series by the output process of a Gaussian system.

For the approximation problem of single-output processes the least-squares criterion and the likelihood function are often used and the resulting algorithms have been extensively analyzed and are widely used. For the approximation of multivariable time series there are

*The research of Dr. A.A. Stoorvogel has been made possible by a fellowship of the Royal Netherlands Academy of Sciences and Arts. His cooperation with the second author is supported in part by CWI.

few algorithms of which the subspace algorithm enjoys the most attention. The subspace algorithm is based on stochastic realization theory and on numerical linear algebra. It was first proposed by H. Akaike and subsequently developed by many researchers. Yet there is no proof that the subspace algorithm is the optimal solution to an approximation problem with a specific criterion.

In this paper a fundamental approach is taken to the problem of approximating a stationary Gaussian process by the output of a Gaussian system. Both processes are related to probability measures. The information theoretic concept of divergence of two probability measures is used as approximation criterion. Divergence is identical to the Kullback-Leibler measure from statistics and probability theory. The approximation problem is then to determine a Gaussian system in a class of systems of at most a prespecified order such that the divergence between the probability measure associated with the output process of this system and the probability measure associated with the given process is as small as possible. This problem was formulated by the second author in [15] and restated in [12].

Two approximation problems are discussed in this paper. The first approximation problem concerns a pair of finite-dimensional Gaussian random variables. This problem is motivated by the main approximation step of the subspace algorithm. The optimal approximation is to perform a transformation to canonical variable form and then truncating the canonical variables to the required order. This result establishes optimality of the main step of the subspace algorithm. The second approximation problem concerns the approximation of a stationary Gaussian process. This problem is currently under investigation. A brief problem formulation is presented in this paper.

A description of the content by section follows. Terminology and notation is stated in Section 2. In Section 3 the approximation problem for finite-dimensional Gaussian random variables is formulated and solved. The approximation problem for stationary Gaussian processes is treated in Section 4.

2 Preliminaries

In this section concepts from analysis and probability are introduced.

General mathematics notation follows. The set of integers is denoted by Z , the set of positive integers by Z_+ , and the set of the natural numbers by N . The set of real numbers is denoted by R and the set of the positive real numbers by $R_+ = [0, \infty)$. The n -dimensional vector space over R is denoted by R^n and the set of $n \times m$ matrices over this vector space by $R^{n \times m}$. The transpose of a matrix $A \in R^{n \times n}$ is denoted by A^T . The matrix $A \in R^{n \times n}$ is said to be positive definite if A is symmetric and $x^T A x \geq 0$ for all $x \in R^n$ and strictly positive definite if A is symmetric and $x^T A x > 0$ for all $x \in R^n, x \neq 0$.

A measurable space, denoted by (Ω, F) , is defined as a set Ω and a σ -algebra F . A probability space is defined as a triple (Ω, F, P) where (Ω, F) is a measurable space and $P : F \rightarrow R$ is a probability measure.

Let $I_A : \Omega \rightarrow R$ be the indicator function of the event $A \in F$ defined by $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ otherwise. For any random variable x let F^x denote the smallest σ -algebra on which the random variable x is measurable.

The sub- σ -algebras F_1 and F_2 of F are said to be *conditionally independent* given the sub- σ -algebra G of F if

$$E[I_{A_1} I_{A_2} | G] = E[I_{A_1} | G] E[I_{A_2} | G],$$

for all $A_1 \in F_1$ and $A_2 \in F_2$. Denote this property by $(F_1, F_2 | G) \in CI$.

Let P, Q be two probability measures on a measurable space (Ω, F) . Then P is said to be absolutely continuous with respect to Q , denoted by $P \ll Q$, if $P(A) = 0$ is implied by $Q(A) = 0$. If $P \ll Q$ then it follows from the Radon-Nikodym theorem that there exists a random variable $r : \Omega \rightarrow R_+$ such that

$$P(A) = E_Q[rI_A] = \int r(\omega)I_A(\omega)dQ.$$

Definition 2.1 Given a measurable space (Ω, F) . Let

$$\mathbf{P} = \{P : F \rightarrow R_+ \mid P \text{ is a probability measure}\}.$$

Define the divergence or the Kullback-Leibler pseudo-distance on \mathbf{P} as $D : \mathbf{P} \times \mathbf{P} \rightarrow R$

$$D(P_1 \parallel P_2) = E_Q \left[r_1 \ln \left(\frac{r_1}{r_2} \right) I_{(r_2 > 0)} \right] = E_{P_1} \left[\ln \left(\frac{r_1}{r_2} \right) I_{(r_2 > 0)} \right],$$

where Q is a σ -finite measure on (Ω, F) , such that $P_1 \ll Q$ and $P_2 \ll Q$, with

$$r_1 = \frac{dP_1}{dQ}, \quad r_2 = \frac{dP_2}{dQ}.$$

In general $D(P_1 \parallel P_2) \neq D(P_2 \parallel P_1)$.

3 Approximation problem of a Gaussian measure on a finite-dimensional space

3.1 Gaussian random variables and divergence

The Gaussian probability distribution function with parameters $m \in R^n$ and $Q \in R^{n \times n}$, with Q strictly positive definite, is defined by the probability density function

$$p(v) = \frac{1}{\sqrt{(2\pi)^n \det Q}} \exp \left(-\frac{1}{2}(v - m)^T Q^{-1}(v - m) \right). \quad (1)$$

A random variable $x : \Omega \rightarrow R^n$ is said to be a Gaussian random variable with parameters $m \in R^n$ and $Q \in R^{n \times n}$, $Q = Q^T$ and Q positive definite, if for all $u \in R^n$

$$E[\exp(iu^T x)] = \exp \left(iu^T m - \frac{1}{2}u^T Q u \right). \quad (2)$$

The notation $x \in G(m, Q)$ will be used in this case. Moreover, $(x_1, \dots, x_n) \in G(m, Q)$ denotes that, with $x = (x_1, \dots, x_n)^T$, $x \in G(m, Q)$. In this case x_1, \dots, x_n are said to be jointly Gaussian random variables.

A random variable with Gaussian probability distribution function is a Gaussian random variable. A Gaussian random variable does not necessarily have a Gaussian probability density function, only so in the case its variance is strictly positive definite.

For the approximation problem of Gaussian random variables formulated below the geometric approach to Gaussian random variables is followed. In this approach one considers the σ -algebra that a random variable generates rather than the random variable itself. Below only the canonical variable form of the geometric approach is defined.

Definition 3.1 Let $y_1 : \Omega \rightarrow R^{k_1}$ and $y_2 : \Omega \rightarrow R^{k_2}$ be jointly Gaussian random variables with $(y_1, y_2) \in G(0, Q)$. Then (y_1, y_2) are said to be in canonical variable form if

$$Q = \begin{pmatrix} I & L \\ L & I \end{pmatrix} \in R^{(k_1+k_2) \times (k_1+k_2)}, L = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \in R^{k_1 \times k_2},$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{k_{12}}) \in R^{k_{12} \times k_{12}}, \quad k_{12} \in N, \quad 1 \geq \lambda_1 \geq \dots \geq \lambda_{k_{12}} > 0.$$

One then says that the components of y_1 and y_2 , $(y_{11}, \dots, y_{1k_1})$ and $(y_{21}, \dots, y_{2k_2})$, are the canonical variables and $(\lambda_1, \dots, \lambda_{k_{12}})$ the canonical correlation coefficients.

Theorem 3.2 [4, 11]. Let $y_1 : \Omega \rightarrow R^{k_1}$ and $y_2 : \Omega \rightarrow R^{k_2}$ be jointly Gaussian random variables with $(y_1, y_2) \in G(0, Q)$. Then there exists a basis transformation $S = \text{Block-diag.}(S_1, S_2)$ such that with respect to the new basis $(S_1 y_1, S_2 y_2) \in G(0, Q_1)$ has the canonical variable form.

The transformation to canonical variable form is not unique in general. Uniqueness occurs only if the canonical correlation coefficients are all different and $k_{12} = k_1 = k_2$. The remaining invariance of the canonical variable form will not be stated here because of space limitation. For additional results on canonical variables the reader is referred to [4, 5]. Note that $Q > 0$ implies that the canonical correlation coefficients satisfy $1 > \lambda_1$.

The divergence of two Gaussian measures on R^n is stated below.

Proposition 3.3 [12, Appendix D]. Let $G(m_1, Q_1)$ and $G(m_2, Q_2)$ be two Gaussian measures on R^n . Assume that $Q_1 > 0$ and $Q_2 > 0$. The divergence between these measures is given by the formula

$$D(G(m_1, Q_1) \| G(m_2, Q_2))$$

$$= -\ln \left(\frac{\det Q_1}{\det Q_2} \right) + \text{tr}([Q_2^{-1} - Q_1^{-1}]Q_1) + (m_1 - m_2)^T Q_2^{-1} (m_1 - m_2) \quad (3)$$

$$= \sum_{i=1}^n [\lambda_i(Q_1, Q_2) - \ln \lambda_i(Q_1, Q_2) - 1] + (m_1 - m_2)^T Q_2^{-1} (m_1 - m_2), \quad (4)$$

where $\{\lambda_i(Q_1, Q_2), i \in 1, \dots, n\}$ are the generalized eigenvalues of Q_1 with respect to Q_2 , or the solutions of the polynomial equation in λ

$$\det(Q_2 \lambda - Q_1) = 0. \quad (5)$$

3.2 Problem formulation

In stochastic realization theory there has been defined the analogon of a stochastic system for random variables. Consider two Gaussian random variables on a finite-dimensional space y_1, y_2 . For the following discussion the reader may think of the analogy with the subspace identification algorithm. Let then y_1 represent the future of a process on a finite horizon at a given time and y_2 represent the past of the same process on another finite horizon at the same time. A *state* for these random variables is then defined to be a Gaussian random variable x on a finite-dimensional space, say R^n , such that (y_1, y_2, x) are jointly Gaussian variables and

$$(F^{y_1}, F^{y_2} | F^x) \in \text{CI},$$

or the sub- σ -algebra F^x makes the sub- σ -algebras F^{y_1} and F^{y_2} conditionally independent. An example of such a state x is $x = Q_{12}Q_{22}^{-1}y_2$ where $(y_1, y_2) \in G(0, Q)$ with

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}.$$

There are many such state variables, see [14] for the details. The space in which the state variable x takes values, R^n , has minimal dimension iff $n = \text{rank}(Q_{12})$.

Consider a pair of Gaussian random variables $(y_{01}, y_{02}) \in G(0, Q_0)$ with $y_{01} : \Omega \rightarrow R^{p_1}$, $y_{02} : \Omega \rightarrow R^{p_2}$, and

$$Q_0 = \begin{pmatrix} Q_{011} & Q_{012} \\ Q_{012}^T & Q_{022} \end{pmatrix}.$$

As mentioned above, y_{01} , y_{02} may represent the future and the past of a process at a particular time. Generically $\text{rank}(Q_{012})$ will be of the order of $\min(p_1, p_2)$. Consequently the state of these variables has a dimension at least that large. For modeling problems this dimension is often much too high. Realistic modeling therefore leads to the problem of approximating the given measure $G(0, Q_0)$ by another Gaussian measure on the same space, say $G(0, Q)$, with

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}, \quad \text{rank}(Q_{12}) \leq n, \quad n \in N \quad (6)$$

such that the distance between both measures is minimal according to some criterion. In this paper we use as a criterion the divergence between both measures. The approximating measure $G(0, Q)$ will then admit a state variable of dimension n or less, where $n \in N$ can be used as a design variable. This motivates the following problem.

Problem 3.4 Consider a Gaussian measure $G(m_0, Q_0)$ on R^p with $p, p_1, p_2 \in Z_+$, $p_1 + p_2 = p$, $m_0 \in R^p$, and $Q_0 > 0$. Fix $n \in N$. Define the model class

$$\begin{aligned} \mathbf{Q}(p_1, p_2, n) &= \{Q \in R^{p \times p} \mid Q = Q^T > 0, \text{rank}(Q_{12}) \leq n\}, \\ \mathbf{G}(n) &= \{G(m, Q) \text{ on } R^p \mid m \in R^p, Q \in \mathbf{Q}(p_1, p_2, n), p = p_1 + p_2\}, \end{aligned}$$

where $Q \in \mathbf{Q}(p_1, p_2, n)$ has the decomposition

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}, \quad Q_{11} \in R^{p_1 \times p_1}, \quad Q_{12} \in R^{p_1 \times p_2}, \quad Q_{22} \in R^{p_2 \times p_2}. \quad (7)$$

The problem is to solve

$$\inf_{G(m, Q) \in \mathbf{G}(n)} D(G(m_0, Q_0) \| G(m, Q)),$$

or, equivalently,

$$\inf_{Q \in \mathbf{Q}(p_1, p_2, n)} [\text{tr}(Q^{-1}Q_0) - \ln \det(Q^{-1}Q_0) + (m_0 - m)^T Q^{-1}(m_0 - m)].$$

This includes showing existence of a measure which attains the infimum and, if so, establishing whether an infimum is unique and, if not, to classify all infima.

The problem has been motivated above. The divergence criterion is related to the likelihood function as is well known. Apparently H. Akaike in [1, 2, 3] first published about this relation. In this regard, see also [12].

The problem is also motivated by the subspace algorithm for the approximation of stationary Gaussian processes. The nucleus of the subspace identification algorithm is an algorithm for the approximation problem of Gaussian random variables. The latter problem is like Problem 3.4, but the divergence criterion is not used. The subspace algorithm was proposed by H. Akaike, see [2]. It is based on stochastic realization theory. The algorithm was later extended and improved by many researchers, see [8, 10, 9, 13].

Approximation problems for Gaussian random variables have been considered already by C.R. Rao in [11], but apparently not with the decomposition (7) and not with the divergence criterion. The problem above was first formulated by the second author in [15], see also [12]. Mr. M. Stöhr first performed research on the problem at CWI in 1989 under supervision of the second author of this paper. Problem 3.4 is difficult because it is an optimization problem over a space which is not convex. The nonconvexity is due to the rank constraint.

3.3 Optimal approximating measure

Algorithm 3.5 Consider given the Gaussian measure $G(m_0, Q_0)$ on R^p , as in Problem 3.4, with $Q_0 = Q_0^T > 0$.

1. Transform Q_0 to a canonical variable form, say

$$Q_0 = S\bar{Q}_0S^T$$

where

$$S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \in R^{p \times p}, \quad \bar{Q}_0 = \begin{pmatrix} I & L \\ L^T & I \end{pmatrix} \in R^{p \times p}, \quad L = \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \in R^{p_1 \times p_2},$$

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k) \in R^{k \times k}$$

where $k \in N$, $\lambda_1, \dots, \lambda_k \in R_+$, $1 > \lambda_1 \geq \dots \geq \lambda_k > 0$, $S_1 \in R^{p_1 \times p_1}$, and $S_2 \in R^{p_2 \times p_2}$.

2. If $0 \leq n \leq k$ then set

$$\begin{aligned} \Lambda_1 &= \text{diag}(\lambda_1, \dots, \lambda_n) \in R^{n \times n}, \\ \Lambda_2 &= \text{diag}(\lambda_{n+1}, \dots, \lambda_k) \in R^{(k-n) \times (k-n)}, \\ L_1 &= \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix} \in R^{p_1 \times p_2}, \quad S^* = S, \quad Q^* = S^* \begin{pmatrix} I & L_1 \\ L_1^T & I \end{pmatrix} (S^*)^T. \end{aligned}$$

If $k < n$ then set $\Lambda_1 = \Lambda$ in the formula for L_1 .

3. Set $m^* = m_0 \in R^p$.

Theorem 3.6 Consider Problem 3.4. The optimal solution to this problem is the Gaussian measure $G(m^*, Q^*)$ with (m^*, Q^*) as defined in Algorithm 3.5. This solution is unique

except when $n > 0$ and $\lambda_n = \lambda_{n+1}$. In the latter case the set of all solutions is described as follows. Suppose that

$$\lambda_{r_1} > \lambda_{r_1+1} = \dots = \lambda_n = \lambda_{n+1} = \dots = \lambda_{r_2} > \lambda_{r_2+1},$$

for $r_1 \in N$ with $r_1 \leq n - 1$ the smallest integer such that either $\lambda_{r_1} > \lambda_{r_1+1}$ or with $r_1 = 0$, and $r_2 \in Z_+$ with $r_2 \geq n + 1$. Let $T_0 \in R^{(r_2-r_1) \times (r_2-r_1)}$ be an orthogonal matrix ($T_0 T_0^T = I = T_0^T T_0$). Define

$$\begin{aligned} T_1 &= \begin{pmatrix} I_{r_1} & 0 & 0 \\ 0 & T_0 & 0 \\ 0 & 0 & I_{(p_1-r_2)} \end{pmatrix} \in R^{p_1 \times p_1}, \\ T_2 &= \begin{pmatrix} I_{r_1} & 0 & 0 \\ 0 & T_0 & 0 \\ 0 & 0 & I_{(p_2-r_2)} \end{pmatrix} \in R^{p_2 \times p_2}, \\ T &= \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \in R^{p \times p}, \quad S^* = ST \in R^{p \times p}, \\ Q^*(T_0) &= S^* \begin{pmatrix} I & L_1 \\ L_1^T & I \end{pmatrix} (S^*)^T \in R^{p \times p}. \end{aligned}$$

The set of optimal variance matrices is then specified by

$$\{Q^*(T_0) \in R^{p \times p} | T_0 \in R^{(r_2-r_1) \times (r_2-r_1)} \text{ orthogonal}\}. \quad (8)$$

The infimal value is

$$p - \sum_{i=n+1}^k \ln(1 - \lambda_i^2(\Lambda)),$$

where $\{\lambda_i(\Lambda), i = 1, \dots, k\}$ are the decreasingly ordered eigenvalues of Λ which are equal to the canonical correlation coefficients of Q_0 .

The Algorithm 3.5 has been known for a long time, it is already stated in [11]. Novel is here only that it is the optimal solution to the approximation problem with the divergence criterion. The notation $*$ as in Q^* is used to indicate the minimum of an infimization problem.

Proof Because $Q \in \mathbf{Q}(n)$, $Q > 0$, hence $m^* = m_0$. In the following it is supposed that $m_0 = 0$ hence $m^* = 0$. The cases $n = 0$ and $n > 0$ are treated successively in the proof.

1. For $n = 0$, $Q \in \mathbf{Q}(p_1, p_2, n)$ has the block-diagonal form

$$Q = \begin{pmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{pmatrix} \in R^{p \times p}.$$

Because $Q > 0$, $P = Q^{-1}$ exists and is also an element of $\mathbf{Q}(p_1, p_2, 0)$ and hence is also block-diagonal. The problem is then equivalent to

$$\inf_{P \in \mathbf{Q}(p_1, p_2, 0)} [\text{tr}(PQ_0) - \ln \det(PQ_0)].$$

2. Transform Q_0 to a canonical variable form as indicated in Step 1 of Algorithm 3.5. Define $\bar{P} = S^T P S$. Because P and S are block-diagonal of compatible sizes, $\bar{P} \in \mathbf{Q}(p_1, p_2, 0)$

is also block-diagonal. Because for square matrices A and B , $\text{tr}(AB) = \text{tr}(BA)$ and $\det(AB) = \det(BA)$, the cost function becomes

$$\text{tr}(\bar{P}\bar{Q}_0) - \ln \det(\bar{P}\bar{Q}_0),$$

where

$$\bar{Q}_0 = \begin{pmatrix} I & L \\ L^T & I \end{pmatrix}.$$

Since, by assumption, $Q_0 > 0$, we find $1 > \lambda_1 \geq \dots \geq \lambda_k > 0$.

3. It follows from the canonical variable form, see Step 1 of Algorithm 3.5 that the underlying variables y_1, y_2 can be decomposed into components as (y_{11}, y_{12}) and (y_{21}, y_{22}) , where y_{11}, y_{12} are independent, y_{21}, y_{22} are independent, and (y_{11}, y_{21}) taken together are independent of (y_{12}, y_{22}) taken together. The approximation problem for y_1, y_2 can then be decomposed into one for (y_{11}, y_{21}) and one for (y_{12}, y_{22}) . In this step this decomposition is formally established. Let

$$T = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \in R^{p \times p}$$

where the successive diagonal blocks of T take values in $R^{k \times k}$, $R^{(p_1-k) \times (p_1-k)}$, $R^{k \times k}$, $R^{(p_2-k) \times (p_2-k)}$, respectively and the other blocks conformly. Then $T^{-1} = T^T$. Decompose \bar{P} and \bar{Q}_0 conform to T as

$$\bar{P} = \begin{pmatrix} P_5 & P_9 & 0 & 0 \\ P_9^T & P_6 & 0 & 0 \\ 0 & 0 & P_7 & P_{10} \\ 0 & 0 & P_{10}^T & P_8 \end{pmatrix}, \quad \bar{Q}_0 = \begin{pmatrix} I & 0 & \Lambda & 0 \\ 0 & I & 0 & 0 \\ \Lambda & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}.$$

Then

$$\begin{aligned} \bar{P}\bar{Q}_0 &= T^T \begin{pmatrix} P_5 & 0 & P_9 & 0 \\ 0 & P_7 & 0 & P_{10} \\ P_9^T & 0 & P_6 & 0 \\ 0 & P_{10}^T & 0 & P_8 \end{pmatrix} \begin{pmatrix} I & \Lambda & 0 & 0 \\ \Lambda & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix} T \\ &= T^T \begin{pmatrix} P_3 & P_{11} \\ P_{11}^T & P_4 \end{pmatrix} \begin{pmatrix} Q_3 & 0 \\ 0 & I \end{pmatrix} T, \end{aligned}$$

where

$$\begin{aligned} P_3 &= \begin{pmatrix} P_5 & 0 \\ 0 & P_7 \end{pmatrix} \in R^{2k \times 2k}, \quad P_4 = \begin{pmatrix} P_6 & 0 \\ 0 & P_8 \end{pmatrix} \in R^{(p-2k) \times (p-2k)}, \\ Q_3 &= \begin{pmatrix} I & \Lambda \\ \Lambda & I \end{pmatrix} \in R^{2k \times 2k}, \quad P_{11} = \begin{pmatrix} P_9 & 0 \\ 0 & P_{10} \end{pmatrix} \in R^{2k \times (p-2k)}. \end{aligned}$$

Then

$$\begin{aligned} \text{tr}(\bar{P}\bar{Q}_0) &= \text{tr} \left(\begin{pmatrix} P_3 & P_{11} \\ P_{11}^T & P_4 \end{pmatrix} \begin{pmatrix} Q_3 & 0 \\ 0 & I \end{pmatrix} \right) = \text{tr}(P_3 Q_3) + \text{tr}(P_4), \\ \ln \det(\bar{P}\bar{Q}_0) &= \ln \det \begin{pmatrix} P_3 Q_3 & P_{11} \\ P_{11}^T Q_3 & P_4 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= \ln \det \begin{pmatrix} P_3 Q_3 & 0 \\ 0 & P_4 - P_{11}^T Q_3 (P_3 Q_3)^{-1} P_{11} \end{pmatrix} \\
&= \ln \det(P_3 Q_3) + \ln \det(P_4 - P_{11}^T P_3^{-1} P_{11}).
\end{aligned}$$

Note that

$$P_4 - P_{11}^T P_3^{-1} P_{11} \leq P_4, \quad \ln \det(P_4 - P_{11}^T P_3^{-1} P_{11}) \leq \ln \det(P_4).$$

Hence

$$\begin{aligned}
&\inf_{\bar{P}} [\operatorname{tr}(\bar{P} \bar{Q}_0) - \ln \det(\bar{P} \bar{Q}_0)] \\
&= \inf_{P_3, P_4, P_{11}} [\operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3) + \operatorname{tr}(P_4) - \ln \det(P_4 - P_{11}^T P_3^{-1} P_{11})] \\
&\geq \inf_{P_3} [\operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3)] + \inf_{P_4} [\operatorname{tr}(P_4) - \ln \det(P_4)],
\end{aligned}$$

while the left-hand side can attain the right-hand side for $P_{11} = 0$. Thus the problem has been reduced to the two problems

$$\begin{aligned}
&\inf_{P_3 \in \mathbf{Q}(k, k, 0)} [\operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3)], \\
&\inf_{P_4 \in \mathbf{Q}(p_1 - k, p_2 - k, 0)} [\operatorname{tr}(P_4) - \ln \det(P_4)].
\end{aligned}$$

4. The infimization problem for the matrix P_3 is shown to have a unique solution if a solution exists. Define the function

$$g : \mathbf{Q}(k, k, 0) \rightarrow R, \quad g(P_3) = \operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3).$$

The space $\mathbf{Q}(k, k, 0)$ is convex. It can be shown that on this set the function g is strictly convex. Thus if the problem

$$\inf_{P_3 \in \mathbf{Q}(k, k, 0)} [\operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3)] \tag{9}$$

has a solution then the solution is unique.

5. The infimization problem for the matrix P_3 is solved in Step 5 and Step 6. Let

$$T_3 = \begin{pmatrix} 0 & I_k \\ I_k & 0 \end{pmatrix} \in R^{2k \times 2k},$$

where the four blocks are in $R^{k \times k}$. The matrix T is invertible with $T^{-1} = T$. Then

$$\begin{aligned}
T_3 P_3 T_3^{-1} &= T_3 \begin{pmatrix} P_5 & 0 \\ 0 & P_7 \end{pmatrix} T_3^{-1} = \begin{pmatrix} P_7 & 0 \\ 0 & P_5 \end{pmatrix}, \\
T_3 Q_3 T_3^{-1} &= Q_3, \\
\operatorname{tr}(P_3 Q_3) - \ln \det(P_3 Q_3) &= \operatorname{tr}((T_3 P_3 T_3^{-1}) Q_3) - \ln \det((T_3 P_3 T_3^{-1}) Q_3).
\end{aligned}$$

Because a solution to (9), if it exists, is unique, it follows that if P_3 is a solution then so is $T_3 P_3 T_3^{-1}$, hence $P_3 = T_3 P_3 T_3^{-1}$ and $P_5 = P_7$. Let $D_1 \in R^{k \times k}$ be a diagonal matrix with on the diagonal +1 or -1, and

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_1 \end{pmatrix} \in R^{2k \times 2k},$$

Then

$$\begin{aligned} DP_3D^{-1} &= \begin{pmatrix} D_1P_5D_1^{-1} & 0 \\ 0 & D_1P_5D_1^{-1} \end{pmatrix} = DP_3D, \\ DQ_3D^{-1} &= Q_3, \\ \text{tr}(P_3Q_3) - \ln \det(P_3Q_3) &= \text{tr}((DP_3D)Q_3) - \ln \det((DP_3D)Q_3). \end{aligned}$$

Again because of the uniqueness of a solution

$$DP_3D = P_3.$$

Let $i, j \in Z_k$, $i \neq j$, $d_i = -d_j$. Then $(P_3)_{ij} = (DP_3D)_{ij} = -(P_3)_{ij}$, hence $(P_3)_{ij} = 0$ and P_3 is diagonal. Then the problem decomposes, with $\lambda_i \in (0, 1)$ and $a_i > 0$, as

$$\begin{aligned} \text{tr}(P_3Q_3) - \ln \det(P_3Q_3) &= \sum_{i=1}^k \text{tr}(P_3(i)Q_3(i)) - \ln \det(P_3(i)Q_3(i)), \\ Q_3(i) &= \begin{pmatrix} 1 & \lambda_i \\ \lambda_i & 1 \end{pmatrix} \in R^{2 \times 2}, \quad P_3(i) = \begin{pmatrix} a_i & 0 \\ 0 & a_i \end{pmatrix} \in R^{2 \times 2}. \end{aligned}$$

6. The single variable case of the infimization problem for the matrix P_3 is solved next. Consider the problem, for $i \in Z_k$,

$$\inf_{a_i \in (0, \infty)} [\text{tr}(P_3(i)Q_3(i)) - \ln \det(P_3(i)Q_3(i))],$$

with $\lambda_i \in (0, 1)$. The cost function is

$$h(a_i) = \text{tr}(P_3(i)Q_3(i)) - \ln \det(P_3(i)Q_3(i)) = 2a_i - 2 \ln(a_i) - \ln(1 - \lambda_i^2).$$

The minimum is then easily calculated by differentiation. The minimum is $a_i^* = 1$, it is unique, and the infimal value is $2 - \ln(1 - \lambda_i^2)$. The solution to the problem of Equation (9) is then $P_3^* = I$ with as infimal value

$$2k - \sum_{i=1}^k \ln(1 - \lambda_i^2).$$

7. Next the infimization problem for the matrix P_4 is considered

$$P_4 \in \mathbf{Q}_{(p_1-k, p_2-k, 0)} \inf [\text{tr}(P_4) - \ln \det(P_4)] \quad (10)$$

where

$$P_4 = \begin{pmatrix} P_6 & 0 \\ 0 & P_8 \end{pmatrix} = P_4^T > 0.$$

Since $x - \ln x - 1 \geq 0$ it is easy to check (by expressing the criterion in the eigenvalues of P_4) that (10) is larger or equal to $p - 2k$ which is the number of columns of P_4 . This lower bound is achieved for $P_4 = I$ which is hence optimal.

8. The results of the Steps 6 and 7 are then used to compute the solution to the problem of the theorem for the case $n = 0$ according to

$$\bar{P}^* = I, \quad P^* = S^{-T} \bar{P}^* S^{-1} = S^{-T} S^{-1} = (SS^T)^{-1}, \quad Q^* = (P^*)^{-1} = SS^T.$$

The infimal value is

$$2k - \sum_{i=1}^k \ln(1 - \lambda_i^2) + (p - 2k) = p - \sum_{i=1}^k \ln(1 - \lambda_i^2).$$

9. Next the problem for the case $n > 0$ is considered. $Q \in \mathbf{Q}(n)$ has the form

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix}, \quad Q_{12} \in R^{p_1 \times p_2}, \quad \text{rank}(Q_{12}) \leq n.$$

Because $Q > 0$, Q^{-1} exists. Let Q^{-1} be decomposed in blocks compatible with Q above. It follows from Proposition A.1 that $\text{rank}(Q_{12}) \leq n$ iff $\text{rank}((Q^{-1})_{12}) \leq n$. Moreover, this proposition yields a formula for Q^{-1} in terms of the blocks of Q and conversely. The problem has then been reduced to solving

$$\inf_{Q^{-1} \in \mathbf{Q}(p_1, p_2, n)} [\text{tr}(Q^{-1}Q_0) - \ln \det(Q^{-1}Q_0)].$$

10. Transform Q_0 to canonical variable form as in Step 1 of Algorithm 3.5. Define $P = S^T Q^{-1} S$. Because S is block diagonal with nonsingular diagonal blocks, $P \in \mathbf{Q}(p_1, p_2, n)$. The criterion now becomes

$$\inf_{P \in \mathbf{Q}(p_1, p_2, n)} [\text{tr}(P\bar{Q}_0) - \ln \det(P\bar{Q}_0)].$$

Decompose this problem as in Step 3. Below the case is treated in which $p_1 = p_2 = k$ and there are k nonzero canonical correlation coefficients.

11. The matrix P is reparametrized such that the approximation problem is reduced to the case $n = 0$ that has already been solved in Step 8. Let

$$\begin{aligned} \mathbf{Q}_{ABC}(p_1, p_2, n) \\ = \{ (A, B, C) \in R^{p_1 \times p_1} \times R^{p_2 \times p_2} \times R^{p_1 \times p_2} \mid A = A^T > 0, B = B^T > 0, \\ \text{rank}(C) \leq n \}, \end{aligned}$$

$$f : \mathbf{Q}_{ABC}(p_1, p_2, n) \rightarrow \mathbf{Q}(p_1, p_2, n),$$

$$P = f(A, B, C) = \begin{pmatrix} I & C \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & 0 \\ C^T & I \end{pmatrix} = \begin{pmatrix} A + CBC^T & CB \\ BC^T & B \end{pmatrix}.$$

It follows that the map f is well defined and has an inverse

$$f^{-1}(P) = (P_{11} - P_{12}P_{22}^{-1}P_{21}^T, P_{22}, P_{12}P_{22}^{-1}).$$

The infimization problem can then be decomposed

$$\begin{aligned} \inf_{(A, B, C) \in \mathbf{Q}_{ABC}(p_1, p_2, n)} [\text{tr}(f(A, B, C)\bar{Q}_0) - \ln \det(f(A, B, C)\bar{Q}_0)] = \\ = \inf_{C \in \mathbf{C}(n)} \left[\inf_{(A, B) \in \mathbf{AB}(n)} \text{tr}(f(A, B, C)\bar{Q}_0) - \ln \det(f(A, B, C)\bar{Q}_0) \right], \end{aligned} \quad (11)$$

where the sets $\mathbf{AB}(n)$ and $\mathbf{C}(n)$ are defined conform $\mathbf{Q}_{ABC}(p_1, p_2, n)$. In the following the inner infimization is considered first in which $C \in \mathbf{C}(n)$ is considered fixed.

12. For the inner infimization of (11) the cost function is now transformed. Let

$$\begin{aligned} \text{tr}(P\bar{Q}_0) &= \text{tr}(f(A, B, C)\bar{Q}_0) = \text{tr} \left(\begin{pmatrix} I & C \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & 0 \\ C^T & I \end{pmatrix} \bar{Q}_0 \right) \\ &= \text{tr} \left(\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & C + \Lambda \\ C^T + \Lambda & H \end{pmatrix} \right), \end{aligned}$$

where

$$H = I + C^T \Lambda + \Lambda C + C^T C.$$

Let $H^{1/2}$ be a symmetric square root of H . Because $\bar{Q}_0 > 0$, $H^{1/2}$ is nonsingular. Let $\bar{B} = H^{1/2} B H^{1/2}$. Then

$$\begin{aligned} \text{tr}(P\bar{Q}_0) &= \text{tr} \left(\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} I & C + \Lambda \\ C^T + \Lambda & H \end{pmatrix} \right) \\ &= \text{tr} \left(\begin{pmatrix} A & 0 \\ 0 & \bar{B} \end{pmatrix} \begin{pmatrix} I & (C + \Lambda)H^{-1/2} \\ H^{-1/2}(C^T + \Lambda) & I \end{pmatrix} \right). \end{aligned} \quad (12)$$

Transform the second matrix in the trace of (12) to canonical variable form,

$$\begin{aligned} H^{-1/2}(C^T + \Lambda) &= U \bar{L}^T V, \quad U \in R^{p_2 \times p_2}, \quad V \in R^{p_1 \times p_1}, \quad \bar{L} \in R^{p_1 \times p_2}, \\ \tilde{Q}_0 &= \begin{pmatrix} I & \bar{L} \\ \bar{L}^T & I \end{pmatrix}, \quad \tilde{A} = V A V^T, \quad \tilde{B} = U^T \bar{B} U, \\ \text{tr}(P\bar{Q}_0) &= \text{tr} \left(\begin{pmatrix} \tilde{A} & 0 \\ 0 & \tilde{B} \end{pmatrix} \tilde{Q}_0 \right). \end{aligned}$$

The inner infimization problem has then been reduced to

$$\inf_{(\tilde{A}, \tilde{B}) \in \mathbf{AB}(n)} \left[\text{tr} \left(\begin{pmatrix} \tilde{A} & 0 \\ 0 & \tilde{B} \end{pmatrix} \tilde{Q}_0 \right) - \ln \det \left(\begin{pmatrix} \tilde{A} & 0 \\ 0 & \tilde{B} \end{pmatrix} \tilde{Q}_0 \right) \right].$$

The solution to the latter problem is provided in Step 6 of the proof. The optimal solution is $(\tilde{A}^*, \tilde{B}^*) = (I, I)$ and the infimal value is

$$2k - \sum_{i=1}^{k_1} \ln(1 - \tilde{\lambda}_i^2),$$

where $\{\tilde{\lambda}_i, i = 1, \dots, k_1\}$ are the canonical correlation coefficients of \tilde{Q}_0 and $k_1 \in N$ is the number of nonzero canonical correlation coefficients. Then $A^* = I$, $\bar{B}^* = I$, and $B^* = H^{-1}$. The inner infimization is then solved.

13. For the outer infimization in (11) an expression is needed of the value of the criterion of Step 12 in terms of the matrix C over which one has to infimize. This value may be deduced from $\{\tilde{\lambda}_i, i = 1, \dots, k_1\}$, the canonical values of \tilde{Q}_0 , or, equivalently, the singular values of $H^{-1/2}(C^T + \Lambda)$. From the matrix inversion lemma follows that

$$\begin{aligned} (C + \Lambda)[I + C^T \Lambda + \Lambda C + C^T C]^{-1}(C^T + \Lambda) &= \\ &= (C + \Lambda)[(I - \Lambda^2) + (C^T + \Lambda)(C + \Lambda)]^{-1}(C^T + \Lambda) \\ &= I - [I + (C + \Lambda)(I - \Lambda^2)(C^T + \Lambda)]^{-1}. \end{aligned}$$

Let $\bar{C}^T = (I - \Lambda^2)^{-1/2} C$, $\tilde{\Lambda} = (I - \Lambda^2)^{-1/2} \Lambda$. Denote the singular values of the matrix \bar{C} by $\{\lambda_i(\bar{C}), i \in Z_k\}$. Thus the singular values of $H^{-1/2}(C^T + \Lambda)$ satisfy

$$\begin{aligned} \lambda_i^2(H^{-1/2}(C^T + \Lambda)) &= 1 - [1 + \lambda_i^2(\bar{C}^T + \tilde{\Lambda})]^{-1}, \\ 1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda)) &= [1 + \lambda_i^2(\bar{C}^T + \tilde{\Lambda})]^{-1}. \end{aligned}$$

In the following consider the decreasingly ordered singular values of \bar{C} and $\tilde{\Lambda}$, $\{\lambda_i(\bar{C}), i \in Z_k\}$ and $\{\lambda_i(\tilde{\Lambda}), i \in Z_k\}$. It follows from [7, Th. 3.3.16] and from $\tilde{\Lambda} = (\bar{C} + \tilde{\Lambda}) + (-\bar{C})$ that

$$\lambda_{i+j-1}(\tilde{\Lambda}) \leq \lambda_i(\bar{C}^T + \tilde{\Lambda}) + \lambda_j(-\bar{C}^T),$$

for all $i, j \in Z_k$ with $i + j \leq k + 1$. Because $\text{rank}(\bar{C}) = \text{rank}(C) \leq n$, $\lambda_{n+1}(\bar{C}^T) = 0$, hence

$$\lambda_{i+n}(\tilde{\Lambda}) \leq \lambda_i(\bar{C}^T + \tilde{\Lambda}),$$

for $i = 1, \dots, k - n$. Note that

$$\begin{aligned} \tilde{\Lambda}^2 &= \Lambda(I - \Lambda^2)^{-1}\Lambda = -I + (I - \Lambda^2)^{-1}, \\ [1 + \lambda_i(\tilde{\Lambda})^2]^{-1} &= 1 - \lambda_i^2(\Lambda). \end{aligned}$$

Then, for $i = 1, \dots, k - n$,

$$1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda)) = [1 + \lambda_i^2(\bar{C}^T + \tilde{\Lambda})]^{-1} \leq [1 + \lambda_i^2(\tilde{\Lambda})]^{-1}.$$

Note also that for $i = 1, \dots, n$

$$1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda)) \leq 1.$$

14. The outer infimization of (11) is then

$$\inf_{C \in \mathbf{C}(n)} \left[2k - \sum_{i=1}^{k_1} \ln(1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda))) \right],$$

where $C \in \mathbf{C}(n)$ is such that $C \in R^{p_1 \times p_2}$ and $\text{rank}(C) \leq n$. Because the logarithm is an increasing function the problem is equivalent to

$$\sup_{C \in \mathbf{C}(n)} \prod_{i=1}^{k_1} [1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda))].$$

An upper bound for the cost function now follows from Step 13 of the proof,

$$\prod_{i=1}^{k_1} [1 - \lambda_i^2(H^{-1/2}(C^T + \Lambda))] \leq \prod_{i=n+1}^{k_1} [1 + \lambda_i^2(\tilde{\Lambda})]^{-1} = \prod_{i=n+1}^{k_1} [1 - \lambda_i^2(\Lambda)],$$

and the right-hand side does not depend on $C \in \mathbf{C}(n)$. Next it is shown that the upper bound can be attained for

$$C^* = \begin{pmatrix} -\Lambda_1 & 0 \\ 0 & 0 \end{pmatrix},$$

where Λ_1 is defined in Step 2 of Algorithm 3.5. Note that there may be more than one matrix C^* which attains the right-hand side in case $\lambda_n = \lambda_{n+1}$. The equivalence class of optimal C^* matrices can be shown to correspond to the set of solutions described in the statement of the theorem. Then

$$H = \begin{pmatrix} I - \Lambda_1^2 & 0 \\ 0 & I \end{pmatrix}, \quad (C^* + \Lambda)H^{-1}((C^*)^T + \Lambda) = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_2^2 \end{pmatrix},$$

$$\lambda_i(H^{-1/2}((C^*)^T + \Lambda)) = \begin{cases} 0, & i = 1, \dots, n, \\ \lambda_i(\Lambda), & i = n + 1, \dots, k, \end{cases}$$

$$\prod_{i=1}^k [1 - \lambda_i^2(H^{-1/2}((C^*)^T + \Lambda))] = \prod_{i=n+1}^k [1 - \lambda_i^2(\Lambda)].$$

15. With the result of Steps 12 and 14 the optimal solution is then calculated as

$$\begin{aligned}
P^* &= f(A^*, B^*, C^*) = \begin{pmatrix} I + \Lambda_1^2(I - \Lambda_1^2)^{-1} & 0 & -\Lambda_1(I - \Lambda_1^2)^{-1} & 0 \\ 0 & I & 0 & 0 \\ -\Lambda_1(I - \Lambda_1^2)^{-1} & 0 & (I - \Lambda_1^2)^{-1} & 0 \\ 0 & 0 & 0 & I \end{pmatrix}, \\
(P^*)^{-1} &= \begin{pmatrix} I & 0 & \Lambda_1 & 0 \\ 0 & I & 0 & 0 \\ \Lambda_1 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix} = \begin{pmatrix} I & L_1 \\ L_1^T & I \end{pmatrix}, \\
Q^* &= S(P^*)^{-1}S^T.
\end{aligned}$$

The minimum value of the criterion is

$$(p - 2k) + 2k - \sum_{i=n+1}^k \ln(1 - \lambda_i^2(\Lambda)) = p - \sum_{i=n+1}^k \ln(1 - \lambda_i^2(\Lambda)).$$

□

4 Divergence approximation of a stationary Gaussian process

Consider a discrete-time stochastic process $y : \Omega \times T \rightarrow R^p$ on $T = Z$. Consider two probability measures for this process, P_0 and P_1 . With respect to P_0 y is stationary, Gaussian, has zero mean value function, and covariance function $W_0 : T \rightarrow R^{p \times p}$, while with respect to P_1 y is stationary, Gaussian, has also zero mean value function, and covariance function $W_1 : T \rightarrow R^{p \times p}$.

The divergence rate between P_0 and P_1 is defined by the formula

$$D_r(P_0 \| P_1) = \lim_{n \rightarrow \infty} \frac{1}{2n+1} D(P_0|_{[-n, +n]} \| P_1|_{[-n, +n]}),$$

if the limit exists, where $P_0|_{[-n, +n]}$, $P_1|_{[-n, +n]}$ denote the restrictions of P_0 and P_1 respectively to a stochastic process on the time index set $\{-n, \dots, -1, 0, 1, \dots, n\}$. Suppose that the process y with respect to P_0 and P_1 admits spectral densities $\hat{W}_0, \hat{W}_1 : C \rightarrow C^{p \times p}$ respectively. The divergence rate between the measures induced by the processes exists and is given by the formula

$$\begin{aligned}
D_r(P_0 \| P_1) &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \left(\frac{\det \hat{W}_1(e^{i\lambda})}{\det \hat{W}_0(e^{i\lambda})} \right) + \text{tr} \left(\hat{W}_1(e^{i\lambda})^{-1} [\hat{W}_0(e^{i\lambda}) - \hat{W}_1(e^{i\lambda})] \right) d\lambda.
\end{aligned}$$

For details on this see [12, Appendix E].

Consider a *Gaussian system* on $T = Z$ and on the probability space (Ω, F) represented by the recursion

$$\begin{aligned}
x(t+1) &= Ax(t) + Bv(t), \\
y(t) &= Cx(t) + Dv(t),
\end{aligned}$$

where $v : \Omega \times T \rightarrow R^m$ is a Gaussian white noise process with, for $t \in T$, $v(t) \in G(0, V)$, $V = V^T \geq 0$, $x : \Omega \times T \rightarrow R^n$ is a stochastic process called the *state process*, $y : \Omega \times T \rightarrow R^p$ is a stochastic process called the *output process*, and $A \in R^{n \times n}$, $C \in R^{p \times n}$, $M \in R^{n \times m}$,

$N \in \mathbb{R}^{p \times m}$. Assume that the eigenvalues of A are strictly inside the unit circle, which property is denoted by $sp(A) \subset C^-$. Then x and y are stationary Gaussian processes.

Call A, C, M, N, V the *parameters* of the Gaussian system and denote the parameter set by

$$\Theta_n = \left\{ \begin{array}{l} (A, C, M, N, V) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{p \times n} \times \mathbb{R}^{n \times m} \times \mathbb{R}^{p \times m} \times \mathbb{R}^{m \times n} \\ V = V^T \geq 0, \quad sp(A) \subset C^- \end{array} \right\}.$$

Call $n \in \mathbb{N}$ the *order* of the Gaussian system.

Denote the set of measures on the output process of a stationary Gaussian system of order less than or equal to $n \in \mathbb{N}$ by

$$\mathbf{P}_{GS(n)} = \{P(\theta), \theta \in \Theta_n\}.$$

With an observed time series one may relate a probability measure P_0 such that with respect to this measure there exists a stationary Gaussian process. This is usually done by taking as the mean value function zero and as covariance function an estimate of the covariance function computed from the data. Not discussed here is the procedure by which one goes from a covariance estimate on a finite horizon to a covariance estimate on the infinite horizon.

Problem 4.1 *Consider a discrete-time Gaussian stochastic process $y : \Omega \times T \rightarrow \mathbb{R}^p$ on $T = \mathbb{Z}$ and a probability space (Ω, F) . Consider a probability measure P_0 on (Ω, F) such that with respect to P_0 y is a stationary Gaussian process with zero mean value function, covariance function $W_0 : T \rightarrow \mathbb{R}^{p \times p}$, and spectral density $\hat{W}_0 : C \rightarrow C^{p \times p}$. The problem is to solve, for fixed $n \in \mathbb{N}$,*

$$\inf_{P(\theta) \in \mathbf{P}_{GS(n)}} D_r(P_0 \| P(\theta)). \quad (13)$$

In principle the infimization of Problem 4.1 can be performed numerically as is done in the maximum likelihood method. The interest of the problem is primarily to show that it can partly be solved analytically and to relate the solution to the subspace algorithm. The approach to Problem 4.1 based on Theorem 3.6 and a limit argument is not correct for several reasons including the fact that the state should also make the past and the future of the state process conditional independent. The authors are currently investigating this problem.

5 Conclusions

Model reduction based on truncation after transformation in a suitable basis is a widely used technique. However, in several cases there is little theoretical foundation for this approximation technique. In [6] model approximation for finite-dimensional linear systems in the Hankel norm was studied in detail. It was shown that truncation in a particular canonical form yields an optimal approximation in the Hankel norm. This paper is a first attempt to perform a similar technique for Gaussian processes. It is shown that optimal approximation using the divergence criterion amounts to truncation in a particular form. The extension of this problem to Gaussian processes is currently being investigated.

Acknowledgements

Support by J.M. van den Hof in the preparation of the paper is gratefully acknowledged.

A Technical results

In this appendix a technical result is stated that is used in the main body of the paper.

Proposition A.1 *Let*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \in R^{(n_1+n_2) \times (n_1+n_2)}$$

with $A_{11} \in R^{n_1 \times n_1}$, $A_{12} \in R^{n_1 \times n_2}$, $A_{21} \in R^{n_2 \times n_1}$, $A_{22} \in R^{n_2 \times n_2}$. Assume that A_{11} , A_{22} , and A are nonsingular.

a Then the matrices $A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $A_{22} - A_{21}A_{11}^{-1}A_{12}$ are nonsingular.

b
$$A^{-1} = \begin{pmatrix} [A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & -A_{11}^{-1}A_{12}[A_{22} - A_{21}A_{11}^{-1}A_{12}]^{-1} \\ -A_{22}^{-1}A_{21}[A_{11} - A_{12}A_{22}^{-1}A_{21}]^{-1} & [A_{22} - A_{21}A_{11}^{-1}A_{12}]^{-1} \end{pmatrix}.$$

References

- [1] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki, editors, *Proceedings 2nd International Symposium Information Theory*, pages 267–281. Akademia Kiado, Budapest, 1973.
- [2] H. Akaike. Canonical correlation analysis of time series and the use of an information criterion. In R.K. Mehra and D.G. Lainiotis, editors, *System identification - Advances and case studies*, pages 27–96. Academic Press, New York, 1976.
- [3] H. Akaike. On entropy maximization principle. In P.R. Krishnaiah, editor, *Applications of statistics*, pages 27–41. North-Holland Publ. Co., Amsterdam, 1977.
- [4] T.W. Anderson. *An introduction to multivariate statistical analysis*. Wiley, New York, 1958.
- [5] R.D. Gittens. *Canonical analysis - A review with applications in ecology*. Springer-Verlag, Berlin, 1985.
- [6] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ -error bounds. *Int. J. Control*, 39:1115–1193, 1984.
- [7] R.A. Horn and C.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.
- [8] A. Lindquist and G. Picci. Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. Report TRITA/MAT-94-32, Department of Mathematics, Royal Institute of Technology, Stockholm, 1995.
- [9] A. Lindquist and G. Picci. Geometric methods for state space identification. In S. Bitanti and G. Picci, editors, *Identification, adaption, learning*, pages 1–69. Springer, London, 1996.
- [10] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.
- [11] C.R. Rao. The use and interpretation of principal component analysis in applied research. *Sankhya, Series A*, 26:329–358, 1964.

- [12] A.A. Stoorvogel and J.H. van Schuppen. System identification with information theoretic criteria. In S. Bittanti and G. Picci, editors, *Identification, adaptation, learning*, pages 289–338. Springer, Berlin, 1996.
- [13] P. van Overschee. *Subspace identification*. PhD thesis, Katholieke Universiteit Leuven, Leuven, 1995.
- [14] C. van Putten and J.H. van Schuppen. The weak and strong gaussian probabilistic realization problem. *J. Multivariate Anal.*, 13:118–137, 1983.
- [15] J.H. van Schuppen. Stochastic realization problems. In J.M. Schumacher H. Nijmeijer, editor, *Three decades of mathematical system theory*, volume 135 of *Lecture Notes in Control and Information Sciences*, pages 480–523. Springer-Verlag, Berlin, 1989.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Preliminaries | 2 |
| 3 | Approximation problem of a Gaussian measure on a finite-dimensional space | 3 |
| 3.1 | Gaussian random variables and divergence | 3 |
| 3.2 | Problem formulation | 4 |
| 3.3 | Optimal approximating measure | 6 |
| 4 | Divergence approximation of a stationary Gaussian process | 14 |
| 5 | Conclusions | 15 |
| A | Technical results | 16 |