Systems & Control Letters 47 (2002) 221 – 226

SYSTEMS & CONTROL LETTERS

www.elsevier.com/locate/sysconle

The Fisher information matrix for linear systems

Raimund J. Ober[a,b,*]

[a] Center for Systems, Communications and Signal Processing, Eric Jonsson School of Electrical Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688, USA

[b] Center for Cancer Immunobiology NB9.106, UT Southwestern Medical Center, 6000 Harry Hines Boulevard, Dallas, TX 75235-8576, USA

Received 15 November 2001; received in revised form 20 April 2002

Abstract

Estimation of parameters of linear systems is a problem often encountered in applications. The Cramer Rao lower bound gives a lower bound on the variance of any unbiased parameter estimation method and therefore provides an important tool in the assessment of a parameter estimation method and for experimental design. Here we study the calculation of the Fisher information matrix, the inverse of the Cramer Rao lower bound, from a system theoretic point of view. A number of results appear in the literature that deal with the case where the stationary data is given as the output of a linear system driven by Gaussian noise. The non-stationary situation where the data is the output of a linear system with Gaussian measurement noise is rarely considered despite its importance in applications. A general description will be given for Fisher information for such data in terms of a derivative system. For a uniformly sampled data set of impulse response type a closed form expression can be given for the Fisher information using the solution of a Lyapunov equation.

Keywords: Fisher information matrix; Cramer Rao lower bound; Lyapunov equation; Linear non-stationary system; Parameter estimation; System identification

1. Introduction

Estimation of parameters that determine dynamic data is a frequently encountered problem in many areas of applications. The question therefore naturally arises as to the accuracy with which these parameters can be estimated. The Cramer Rao lower bound [10,4,2,3] gives a lower bound for the co-variance of the parameter estimates of an unbiased estimation procedure for a given data set. The Cramer Rao lower bound is in fact typically calculated as the inverse of a matrix called the Fisher information matrix. The relevance of the Cramer Rao lower bound is not only to evaluate a particular estimation procedure but it can also give guidance for an appropriate design of an experiment to collect data. In many experimental situations there is a limit on the number of data points that can be acquired. For example, in clinical trials of drugs, patients cannot be subjected to an arbitrary number of blood tests. It is therefore important to choose a sampling strategy that is likely to produce good quality parameter estimates.

Expressions for the Fisher information matrix in system theoretic terms have appeared in the

* Correspondence address: Center for Systems, Communications and Signal Processing, Eric Jonsson School of Electrical Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688, USA.

E-mail address: ober@utdallas.edu (R.J. Ober).

PII: S0167-6911(02)00190-1

literature before in the context of the modelling of stationary time series [11,5,9]. However, there appears to be no systematic investigation of the Cramer Rao lower bound or Fisher information matrix for the case of non-stationary deterministic systems corrupted by measurement noise. For exponentially decaying data, expressions for the Fisher information matrix have appeared in [6]. Similar expressions have, for example, been used in [7] to explore the experimental design for surface plasmon resonance experiments. The contribution of this paper is to systematically analyze the determination of the Fisher information matrix for data that is the output of a deterministic linear system.

Let $(A_\theta, b_\theta, c_\theta, d_\theta)$ be a continuous time single-input single-output system that depends on the parameter vector θ, i.e.

$$\dot{x}_\theta(t) = A_\theta x_\theta(t) + b_\theta u(t), \quad x_\theta(t_0) = x_{\theta,0},$$
$$y_\theta(t) = c_\theta x_\theta(t) + d_\theta u(t), \quad t \geqslant t_0,$$

where A_θ is a $n \times n$ matrix, b_θ is a $n \times 1$ vector, c_θ is a $1 \times n$ vector, d_θ is a scalar, the state vector x_θ is $n \times 1$ and the input u is a typically piecewise continuous function for $t \geqslant t_0$. The parameter space Θ is assumed to be an open subset of $\Re^n$ and the parametrization is assumed to have the appropriate smoothness properties. The output of the system is then given by the standard convolution expression

$$y_\theta(t) = c_\theta e^{(t-t_0)A_\theta} x_{\theta,0} + \int_{t_0}^{t} c_\theta e^{(t-\tau)A_\theta} b_\theta u(\tau)\,d\tau + d_\theta u(t), \quad t \geqslant t_0.$$

We now assume that we have acquired noise corrupted samples $s(k)$, $1 \leqslant k \leqslant N$, of the measured output at various time points t_k, $1 \leqslant k \leqslant N$, with $t_0 \leqslant t_1 < t_2 < \cdots < t_k < \cdots < t_N$, i.e.

$$s_\theta(k) := y_\theta(t_k) + w(k), \quad 1 \leqslant k \leqslant N,$$

where $w(k)$, $k = 1, \ldots, N$, is a sequence of measurement noise which is assumed to be independently normally distributed with mean zero and variance σ_k^2, $k = 1, \ldots, N$. By slight abuse of notation we also allow $N = \infty$ to indicate that $k = 1, \ldots, \infty$.

By the Cramer Rao lower bound [10,4] any unbiased estimator $\hat{\theta}$ of θ has a variance (provided certain regularity conditions hold) such that

$$\mathrm{var}(\hat{\theta}) \geqslant I(\theta)^{-1}.$$

Here $I(\theta)$ is the Fisher information matrix that is given by

$$[I(\theta)]_{st} = -E\left[\frac{\partial^2 \ln(p(S;\theta))}{\partial\theta_s\,\partial\theta_t}\right]_{st},$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_m)^{\mathrm{T}}$, S is the data vector, $p(S;\theta)$ is the probability density function of the measurements and E is the expected value with respect to the underlying probability measure.

An important part of the calculation of the Fisher information matrix for the data model discussed in this paper involves the determination of the derivative the output of the system with respect to the parameters (see Section 2). If an analytical expression is available for the output of a system it can therefore be possible to calculate the Fisher information matrix for a particular system representation. In fact, the results in [6,7] on data that is the linear combination of exponential functions, can be interpreted as data that is given as the transient response of a linear system given in modal form. If the parameterization of the data is changed these tedious calculations typically have to be repeated. For example, if a new system representation/parameterization is given for the data it is not possible to determine the Fisher information matrix unless possibly very extensive derivations are performed. This raises the fundamental question whether there is a system theoretic approach to dealing with the calculation of the Fisher information matrix that reduces these above-mentioned calculations to more standard system theoretic operations.

In Section 2 we show that the calculation of the Fisher information matrix for an arbitrary parameterization of a linear system can be essentially reduced to the calculation of an ℓ^2 type sum of the output of a derivative system of the original system. In the special case where the data is given as a measurement noise corrupted transient response of a linear system it is shown that the calculation of the Fisher information matrix essentially reduces to the calculation of the solution of a Lyapunov equation for the derivative followed by pre- and post-multiplication of the solution by a matrix and its transpose. This means that the calculation of the Fisher information matrix can be reduced to a sequence of well-defined system theoretic operations for which both effective numerical and analytic methods exist (see e.g. [8] for a computer algebra approach to the solution of Lyapunov equations).

We denote by $\mathrm{diag}(M_1, M_2, \ldots, M_r)$ the block diagonal matrix whose diagonal block entries are $M_1, M_2, \ldots, M_r$. All other block entries are zero matrices.

2. Fisher information matrix

In this section we are going to derive an expression for the Fisher information matrix $I(\theta)$ corresponding to the problem that was introduced in the introductory Section 1. If $S = (s_\theta(1), \ldots, s_\theta(N))$ is a data vector note that the probability density function is given by

$$p(S;\theta) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma_k^2}} \mathrm{e}^{-(1/2\sigma_k^2)(s_\theta(k) - y_\theta(t_k))^2}.$$

In the following lemma an immediate result is presented (see also e.g. [6]) and useful notation is introduced.

Lemma 2.1. *With the previous notation and assumptions*

$$I(\theta) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} D_\theta y(t_k) D_\theta y^{\mathrm{T}}(t_k),$$

where for $k = 1, 2, \ldots, N$,

$$D_\theta y(t_k) := \begin{pmatrix} \dfrac{\partial y_\theta}{\partial \theta_1}(t_k) \\ \dfrac{\partial y_\theta}{\partial \theta_2}(t_k) \\ \vdots \\ \dfrac{\partial y_\theta}{\partial \theta_m}(t_k) \end{pmatrix}.$$

Proof. Note that the derivatives of the log-likelihood function with respect to the parameters θ_s and θ_t $1 \leqslant s,\ t \leqslant m$, is given by

$$\frac{\partial^2 \ln(p(S;\theta))}{\partial\theta_s \partial\theta_t} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \left[\left(\frac{\partial^2 y_\theta(t_k)}{\partial\theta_s\,\partial\theta_t} \right) (s(k) - y_\theta(t_k)) - \left(\frac{\partial y_\theta(t_k)}{\partial\theta_s} \right) \left(\frac{\partial y_\theta(t_k)}{\partial\theta_t} \right) \right].$$

The result now follows immediately by taking expectations and using the fact that the residuals $s(k) - y_\theta(t_k)$ have zero mean, $k = 1, \ldots, N$. □

It is now clear that in order to calculate the Fisher information $I(\theta)$ it is necessary to calculate the derivative $\partial y_\theta(t_k)/\partial\theta$ of the system output with respect to the parameter vector θ. To do this we quote a lemma that is essentially the continuous time equivalent of Lemmas 5.2–30 in [2], although we use an alternative method of proof.

Lemma 2.2. *Let* $(A_\theta, b_\theta, c_\theta)$ *be a realization of a function* $h_\theta(t)$, $t \geqslant 0$, *i.e.* $h_\theta(t) = c_\theta \mathrm{e}^{tA_\theta} b_\theta$, $t \geqslant 0$, *where* θ *is a parameter vector. Then for* $1 \leqslant s \leqslant m$ *and* $t \geqslant 0$

$$\frac{\partial h_\theta(t)}{\partial\theta_s} = \partial_s c \mathrm{e}^{t\partial_s A} \partial_s b,$$

where

$$\partial_s c := \begin{bmatrix} \dfrac{\partial c_\theta}{\partial\theta_s} & c_\theta \end{bmatrix}, \quad \partial_s b := \begin{bmatrix} b_\theta \\ \dfrac{\partial b_\theta}{\partial\theta_s} \end{bmatrix}, \quad \partial_s A := \begin{bmatrix} A_\theta & 0 \\ \dfrac{\partial A_\theta}{\partial\theta_s} & A_\theta \end{bmatrix}.$$

Proof. We have for $t \geqslant 0$,

$$\frac{\partial h_\theta(t)}{\partial\theta_s} = \frac{\partial(c_\theta \mathrm{e}^{tA_\theta} b_\theta)}{\partial\theta_s} = \frac{\partial c_\theta}{\partial\theta_s} \mathrm{e}^{tA_\theta} b_\theta + c_\theta \frac{\partial(\mathrm{e}^{tA_\theta} b_\theta)}{\partial\theta_s}$$

and

$$\frac{\partial \mathrm{e}^{tA_\theta} b_\theta}{\partial\theta_s} = \frac{\partial \mathrm{e}^{tA_\theta}}{\partial\theta_s} b_\theta + \mathrm{e}^{tA_\theta} \frac{\partial b_\theta}{\partial\theta_s}.$$

Therefore for $t \geqslant 0$

$$\frac{\partial h_\theta(t)}{\partial\theta_s} = \begin{bmatrix} \dfrac{\partial c_\theta}{\partial\theta_s} & c_\theta \end{bmatrix} \begin{bmatrix} \mathrm{e}^{tA_\theta} & 0 \\ \dfrac{\partial \mathrm{e}^{tA_\theta}}{\partial\theta_s} & \mathrm{e}^{tA_\theta} \end{bmatrix} \begin{bmatrix} b_\theta \\ \dfrac{\partial b_\theta}{\partial\theta_s} \end{bmatrix}.$$

To complete the proof it remains to verify that

$$\begin{bmatrix} \mathrm{e}^{tA_\theta} & 0 \\ \dfrac{\partial \mathrm{e}^{tA_\theta}}{\partial\theta_s} & \mathrm{e}^{tA_\theta} \end{bmatrix} = \mathrm{e}^{t\begin{bmatrix} A_\theta & 0 \\ \partial A_\theta/\partial\theta_s & A_\theta \end{bmatrix}}.$$

To do this first consider for $t \geqslant 0$

$$\mathrm{e}^{t\begin{bmatrix} A_\theta & 0 \\ \partial A_\theta/\partial\theta_s & A_\theta \end{bmatrix}} = \begin{bmatrix} \mathrm{e}^{tA_\theta} & 0 \\ E(t) & \mathrm{e}^{tA_\theta} \end{bmatrix},$$

where the 2,1 entry $E(t)$ is given by

$$E(t) = \sum_{r=0}^{\infty} \frac{(t)^r}{r!} E_r,$$

and the coefficients are inductively seen to be

$$\begin{aligned}
E_0 &= 0,\\
E_1 &= \frac{\partial A_\theta}{\theta_s},\\
E_2 &= \frac{\partial A_\theta}{\theta_s} A_\theta + A_\theta \frac{\partial A_\theta}{\theta_s},\\
E_3 &= \frac{\partial A_\theta}{\theta_s} A_\theta^2 + A_\theta \frac{\partial A_\theta}{\theta_s} A_\theta + A_\theta^2 \frac{\partial A_\theta}{\theta_s},\\
E_4 &= \frac{\partial A_\theta}{\theta_s} A_\theta^3 + A_\theta \frac{\partial A_\theta}{\theta_s} A_\theta^2 + A_\theta^2 \frac{\partial A_\theta}{\theta_s} A_\theta + A_\theta^3 \frac{\partial A_\theta}{\theta_s}.\\
&\vdots
\end{aligned}$$

The 2, 1 entry of

$$\begin{bmatrix} e^{tA_\theta} & 0 \\ \dfrac{\partial e^{tA_\theta}}{\partial \theta_s} & e^{tA_\theta} \end{bmatrix}$$

is given by

$$\begin{aligned}
\frac{\partial e^{tA_\theta}}{\partial \theta_s} &= \frac{\partial \sum_{r=0}^{\infty} t^r (A_\theta)^r / r!}{\partial \theta_s} = \sum_{r=0}^{\infty} \frac{t^r}{r!} \frac{\partial A_\theta^r}{\partial \theta_s}\\
&= \sum_{r=1}^{\infty} \frac{t^r}{r!} \left(\frac{\partial A_\theta}{\partial \theta_s} A_\theta^{r-1} + A_\theta \frac{\partial A_\theta}{\partial \theta_s} A_\theta^{r-2} + \cdots \right.\\
&\qquad \left. + A_\theta^{r-2} \frac{\partial A_\theta}{\partial \theta_s} A_\theta + A_\theta^{r-1} \frac{\partial A_\theta}{\partial \theta_s} \right).
\end{aligned}$$

Comparing coefficients of the expansions of the 2, 1 entries it is verified that

$$\begin{bmatrix} e^{tA_\theta} & 0 \\ \dfrac{\partial e^{tA_\theta}}{\partial \theta_s} & e^{tA_\theta} \end{bmatrix} = e^{t\begin{bmatrix} A_\theta & 0 \\ \partial A_\theta / \partial \theta_s & A_\theta \end{bmatrix}}$$

as remained to be shown. □

In the following theorem we summarize the results and state the general expressions for the Fisher information matrix.

Theorem 2.1. *Consider the augmented derivative system given by*

$$D_\theta A := \mathrm{diag}(\partial_1 A, \partial_2 A, \ldots, \partial_m A),$$

$$D_\theta c := \mathrm{diag}(\partial_1 c, \partial_2 c, \ldots, \partial_m c),$$

$$D_\theta b := \begin{pmatrix} \partial_1 b \\ \partial_2 b \\ \vdots \\ \partial_m b \end{pmatrix}, \quad D_\theta d := \begin{pmatrix} \partial_1 d \\ \partial_2 d \\ \vdots \\ \partial_m d \end{pmatrix}, \quad D_\theta x_0 := \begin{pmatrix} \partial_1 x_0 \\ \partial_2 x_0 \\ \vdots \\ \partial_m x_0 \end{pmatrix}$$

with

$$\partial_s A := \begin{pmatrix} A_\theta & 0 \\ \dfrac{\partial A_\theta}{\partial \theta_s} & A_\theta \end{pmatrix}, \quad \partial_s b := \begin{pmatrix} b_\theta \\ \dfrac{\partial b_\theta}{\partial \theta_s} \end{pmatrix},$$

$$\partial_s c := \begin{pmatrix} \dfrac{\partial c_\theta}{\partial \theta_s} & c_\theta \end{pmatrix}, \quad \partial_s d := \frac{\partial d}{\partial \theta_s},$$

$$\partial_s x_0 := \begin{pmatrix} x_{0,\theta} \\ \dfrac{\partial x_{0,\theta}}{\partial \theta_s} \end{pmatrix}.$$

Then

(1) *for* $t \geqslant t_0$

$$\begin{aligned}
D_\theta y(t) := \begin{pmatrix} \dfrac{\partial y(t)}{\partial \theta_1} \\ \dfrac{\partial y(t)}{\partial \theta_2} \\ \vdots \\ \dfrac{\partial y(t)}{\partial \theta_m} \end{pmatrix}
&= D_\theta c e^{(t-t_0) D_\theta A} D_\theta x_0 \\
&\quad + \int_{t_0}^{t} D_\theta c e^{(t-\tau) D_\theta A} D_\theta b u(\tau)\, d\tau + D_\theta\, du(t);
\end{aligned}$$

(2) *the Fisher information matrix for the noise model introduced in Section* 1 *and the sampling points* $t_1, \ldots, t_N$ *is then given by*

$$I(\theta) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} D_\theta y(t_k) D_\theta^{\mathrm{T}} y(t_k).$$

Proof. (1) The proof follows by differentiating the convolution description of the output $y(t)$ with respect to the individual parameters. Use is made of Lemma 2.2 and the fact that derivation and integration can be exchanged since the integrand is bounded and the integration is over a finite interval. The final expression follows by stacking up the variables for the system.

(2) is the content of Lemma 2.1. □

In the following theorem we are going to consider a special case that is of importance in many applications.

Theorem 2.2. *Assume that the system and data are such that*

(1) *an infinite number of equidistant samples are acquired, i.e.*

$$t_k = (k-1)T, \quad k = 1,2,3,\dots$$

where $T > 0$ is the sampling interval;

(2) *the input u is zero, i.e. $u(t) = 0$, for $t \geqslant t_0$, i.e. the deterministic part of the measured signal is given by*

$$y(t_k) = c\mathrm{e}^{(k-1)TA}x_0, \quad k = 1,2,3,\dots$$

(3) *all the eigenvalues of e^{TA} are in the open unit disk or equivalently the eigenvalues of A are in the open left half plane.*

Then the Fisher information matrix for the noise model described in Section 1, *with $\sigma_k^2 =: \sigma^2$, $k \geqslant 1$, is given by*

$$I(\theta) = \frac{1}{\sigma^2}(D_\theta c)P(D_\theta c)^{\mathrm{T}},$$

where P is the unique solution to the Lyapunov equation

$$A_d P A_d^{\mathrm{T}} - P = -(D_\theta x_0)(D_\theta x_0)^{\mathrm{T}},$$

where $A_d := \mathrm{e}^{TD_\theta A}$.

Proof. Note that e^{TA} is assumed that have all eigenvalues in the open unit disk. By the definition $D_\theta A$ is a block lower triangular matrix with diagonal block entries given by A. Therefore $\mathrm{e}^{TD_\theta A}$ is also block triangular with block diagonal entries given by e^{TA}. Hence $A_d := \mathrm{e}^{TD_\theta A}$ has all eigenvalues in the open unit disk.

With our assumptions the Fisher information matrix is then given by

$$I(\theta) = \frac{1}{\sigma^2}\sum_{k=1}^{\infty} D_\theta y(t_k) D_\theta^{\mathrm{T}} y(t_k).$$

Note that the series converges since $D_\theta y(t_k)$ converges exponentially to zero. We have that

$$\begin{aligned} I(\theta) &= \frac{1}{\sigma^2}\sum_{k=0}^{\infty} D_\theta y(kT) D_\theta^{\mathrm{T}} y(kT) \\ &= \frac{1}{\sigma^2}\sum_{k=0}^{\infty} (D_\theta \mathrm{e}^{kTD_\theta A} D_\theta x_0)(D_\theta \mathrm{e}^{kTD_\theta A} D_\theta x_0)^{\mathrm{T}} \\ &= \frac{1}{\sigma^2} D_\theta c \left(\sum_{k=0}^{\infty} A_d^k D_\theta x_0 D_\theta^{\mathrm{T}} x_0 (A_d^{\mathrm{T}})^k \right) D_\theta^{\mathrm{T}} c \\ &= \frac{1}{\sigma^2} D_\theta c P D_\theta^{\mathrm{T}} c, \end{aligned}$$

where P is the unique solution to the Lyapunov equation

$$A_d P A_d^{\mathrm{T}} - P = -\Delta_\theta x_0 \Delta_\theta^{\mathrm{T}} x_0$$

since A_d has all its eigenvalues in the open unit disk [1]. □

In the above we have discussed the calculation of the Fisher information matrix for data that arises from the output of a sampled continuous time system. If uniform sampling is used, the data can be seen to be the output of a discrete time system. In the following we are going to briefly discuss the results that would apply if the system is in fact discrete time. A discrete time system is given by

$$x_\theta(k+1) = A_\theta x_\theta(k) + b_\theta u(k), \quad x_\theta(0) = x_{\theta,0},$$

$$y_\theta(k) = c_\theta x_\theta(k) + d_\theta u(k), \quad k \geqslant 0,$$

where $u(k)$, $k = 0,\dots,N-1$, is the input sequence. As before we also deal with infinite data sets, i.e. $N = \infty$, is an included possibility. As before we assume that the acquired data $s_\theta(k)$, $0 \leqslant k \leqslant N-1$, is given by

$$s_\theta(k) := y_\theta(k) + w(k),$$

where $w(k)$, $k = 0,\dots,N-1$, is a sequence of zero mean Gaussian random variables with variance σ_k^2, $k = 0,\dots,N-1$. The Fisher information matrix $I(\theta)$, for the m-dimensional parameter vector θ, is calculated analogously to the continuous time case and is given by

$$I(\theta) = \sum_{k=0}^{N-1} \frac{1}{\sigma_k^2} D_\theta y(k) D_\theta y^{\mathrm{T}}(k).$$

With the analogous notation to the continuous-time situation, we have that

$$\begin{aligned} D_\theta y(k) &= D_\theta c (D_\theta A)^k D_\theta x_0 \\ &\quad + \sum_{r=1}^{k} D_\theta c (D_\theta A)^{k-r} D_\theta b u(r-1) + D_\theta\, \mathrm{d}u(k). \end{aligned}$$

We now state the discrete time version of Theorem 2.2 in which it is shown that the Fisher information matrix can be calculated by solving a Lyapunov equation if the data is given by the impulse response of a discrete time system.

Theorem 2.3. *Assume that the original discrete time system is such that*

(1) *the input u is zero, i.e.* $u(k)=0$, *for* $k \geqslant 0$, *i.e. the deterministic part of the measured signal is given by*

$$y(k)=cA^k x_0, \quad k=0,1,2,3,\ldots;$$

(2) *all the eigenvalues of A are in the open unit disk.*

The Fisher information matrix for the noise model described above, with $\sigma_k^2 =: \sigma^2,\ k \geqslant 0$, *is given by*

$$I(\theta)=\frac{1}{\sigma^2}(D_\theta c)P(D_\theta c)^{\mathrm{T}},$$

where P is the unique solution to the Lyapunov equation

$$(D_\theta A)P(D_\theta A)^{\mathrm{T}}-P=-(D_\theta x_0)(D_\theta x_0)^{\mathrm{T}}.$$

Proof. The proof is analogous to the proof of the continuous time version of Theorem 2.2. □

As in the continuous time case the above theorem shows that the calculation of the Fisher information for the case of impulse response data essentially reduces to the calculation of the derivative system, the solution of a Lyapunov equation followed by a simple matrix multiplication. The solution of the Lyapunov equation can be obtained by standard techniques in numerical linear algebra.

Remark 2.1. In [2] a detailed study was carried out in which a Riemannian metric was defined on the space of stable linear systems of fixed dimension. There is an interesting connection between the formula presented above in Theorem 2.3 and those discussed in [2] for the Riemannian metric introduced for the L^2 case. This parallels the better understood connections between the Fisher information for the case of ARMA time series and the associated Riemannian metric.

Acknowledgements

The research was supported in part by grants by the National Science Foundation (DMS 9803186) and the National Institutes of Health (NIH GM58538). The author would like to thank Bernard Hanzon and Zhiping Lin for helpful comments on the manuscript.

References

[1] F.R. Gantmacher, The Theory of Matrices, Chelsea, New York, 1959.
[2] B. Hanzon, Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems: Part 1, CWI Tract, Vol. 63, Centrum voor Wiskunde and Informatica, Amsterdam, The Netherlands, 1980.
[3] R. Johansson, System Modelling and Identification, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[4] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[5] A. Klein, P. Spreij, On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices, Linear Algebra Appl. 237/238 (1996) 579–590.
[6] R. Kumesaran, D.W. Tufts, Estimating the parameters of exponentially damped sinusoids, IEEE Trans. Acoust. Speech Signal Process. 30 (1982) 833–840.
[7] R.J. Ober, E.S. Ward, The influence of signal noise on the accuracy of kinetic constants measured by surface plasmon resonance experiments, Anal. Biochem. 273 (1999) 49–59.
[8] R.L.M. Peeters, B. Hanzon, A Faddeev sequence method for solving Lyapunov and Sylvester equations, Linear Algebra Appl. 241–243 (1996) 401–430.
[9] R.L.M. Peeters, B. Hanzon, Symbolic computation of Fisher information matrices for parametrized state-space systems, Automatica 35 (1999) 1059–1071.
[10] S. Silvey, Statistical Inference, Chapman & Hall, London, 1975.
[11] P. Whittle, The analysis of multiple stationary time series, J. Roy. Statist. Soc. 15 (1953) 125–139.