15th Triennial World Congress, Barcelona, Spain

LOCAL OPTIMALITY OF MINIMUM PHASE BALANCED TRUNCATION

Wolfgang Scherrer [1]

Institut für Ökonometrie, Operations Research und Systemtheorie,
Technische Universität Wien,
Argentinierstr. 8/119, A-1040, Vienna, Austria
W.Scherrer@tuwien.ac.at

Abstract: Balanced model truncation has been considered by many authors, since it is a simple and, nevertheless, efficient model reduction technique. In many cases the approximation error may be bounded by a function of the neglected singular values. In this paper the performance of balanced truncation of state space models for ARMA processes is analysed, where the goodness of fit is measured by the asymptotic Gaussian likelihood function. It is shown that locally, i.e. close to the set of lower order systems, minimum phase balanced truncation and stochastically balanced truncation give almost optimal results.

Keywords: ARMA models, State space models, Model reduction, Likelihood function, Realisation theory, Lyapunov equation

1. INTRODUCTION

Model reduction is concerned with the problem of finding a 'simple' model, which is a good approximation of a 'complex' model. In this paper state space models for discrete time, weakly stationary processes with a rational spectral density, i.e. ARMA processes, are considered. In this setup the asymptotic Gaussian likelihood is a convenient measure of the goodness of fit of the approximate model.

In general, the problem of finding the best k-th order state space model for a process, which is generated by an n-th order system ($n > k$), is a difficult optimisation problem, which only can be solved by numerical optimisation techniques.

On the other hand, balanced model truncation is a simple approach, which gives good results, especially if the true n-th order model is 'close' to the set of all k-th order models.

In this paper, the behaviour of truncation methods, in the case that a sequence of n-th order models converges to a k-th order model will be investigated. It will be shown that, given certain regularity conditions, minimum phase balanced truncation gives the fastest rate of convergence of the truncated model to the best approximation. In this sense, minimum phase balanced truncation is 'locally optimal'.

Note that stochastically balanced realisations are related via diagonal state space transformations to minimum phase balanced realisations. Therefore, stochastically balanced truncation shares this optimality property.

The outline of the paper is as follows: The next section 2 defines the problem setting and gives the main results. Section 3 illustrates the results obtained with a simple simulation example. Finally section 4 gives some conclusions and remarks. Most of the proofs are deferred to section 5.

[1] The author acknowledges support by the Austrian 'Fond zur Förderung der wissenschaftlichen Forschung' through project P-14438INF and by the program 'Training and Mobility of Researchers - Research Networks' through project System Identification (FMRX CT98 0206) and acknowledges contacts with the participants in the European Research Network System Identification (ERNSI).

2. MODEL REDUCTION AND BALANCED MODEL TRUNCATION

Consider a p-dimensional, discrete time, stationary process $(y_t|t \in \mathbb{Z})$ generated by state space system of the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t \end{aligned} \tag{1}$$

where x_t is the n-dimensional state and (u_t) is white noise with variance $\mathrm{E}\, u_t u_t' = \Sigma > 0$. W.l.o.g. it is assumed that $D = I$, and that the system (1) is minimal, stable ($\lambda_{\max}(A) < 1$), and strictly minimum phase ($\lambda_{\max}(A - BC) < 1$); i.e. (1) is in innovation form. Here $\lambda_{\max}(X)$ denotes the maximum modulus of the eigenvalues of a matrix X, i.e. $\lambda_{\max}(X) = \max_i\{|\lambda_i(X)|\}$, where $\lambda_i(X)$ are the eigenvalues of the matrix X.

The model (1) is a realisation of the transfer function

$$\mu(z) = \sum_{i \geq 0} \mu_i z^i = D + C(z^{-1}I - A)^{-1}B \tag{2}$$

where z denotes the backward shift, i.e. $z(y_t|t \in \mathbb{Z}) = (y_{t-1}|t \in \mathbb{Z})$. By minimality, the transfer function has McMillan degree n. In the sequel, $\mathbb{M}(n)$ denotes the set of stable, strictly minimum phase, $(p \times p)$ transfer functions with McMillan degree n.

Throughout this paper a shorthand notation of the form

$$\left.\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right. \tag{3}$$

is used to describe state space systems of the above form. Given the transfer function $\mu(z)$, the system matrices of a minimal realisation are unique only up to transformations of the state space; i.e. for any non singular $T \in \mathbb{R}^{n \times n}$

$$\left.\begin{array}{c|c} \tilde{A} & \tilde{B} \\ \hline \tilde{C} & \tilde{D} \end{array}\right. = \left.\begin{array}{c|c} TAT^{-1} & TB \\ \hline CT^{-1} & D \end{array}\right. \tag{4}$$

is a realisation of $\mu(z)$.

Next, consider a model

$$\left.\begin{array}{c|c} F & G \\ \hline H & I \end{array}\right. \tag{5}$$

for a transfer function $\nu(z) = I + H(z^{-1}I - F)^{-1}G \in \mathbb{M}(k)$, for some $k \leq n$. Together with a variance $\Omega > 0$, this defines an alternative (in general misspecified) model for (y_t). The quality of this model may be assessed by the negative asymptotic Gaussian log likelihood

$$\begin{aligned} l(F, G, H, \Omega; A, B, C, \Sigma) = \\ \log\det\Omega + \mathrm{E}\, v_t' \Omega^{-1} v_t \end{aligned} \tag{6}$$

where (v_t) is the prediction error based on (5):

$$v_t = \nu^{-1}(z) y_t = \nu^{-1}(z)\mu(z) u_t. \tag{7}$$

The minimum value of $l(\cdot)$ is equal to $\log\det\Sigma + p$, which is attained iff $\Omega = \Sigma$ and $\mu(z) = \nu(z)$ holds, i.e. iff $\Omega = \Sigma$ and (5) is equal to (4) for some T.

Model reduction is the problem to find the best model (5) of order $k < n$, with respect to the criterion function (6), i.e. to find parameters (F, G, H, Ω) such that (6) is minimal. In general there is no closed form solution to this optimisation problem, and thus the optimal model has to be found by iterative nonlinear optimisation methods.

On the other hand, principal subsystem truncation (in the following simply called truncation) is a simple approach to get a reduced order model. Let the system matrices be conformingly partitioned as:

$$\left.\begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline C_1 & C_2 & I \end{array}\right. \tag{8}$$

where $A_{11} \in \mathbb{R}^{k \times k}$. Then the truncated model is defined as:

$$\left.\begin{array}{c|c} F & G \\ \hline H & I \end{array}\right. := \left.\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & I \end{array}\right. \tag{9}$$

In general, there is no guarantee that the truncated model is minimal, stable, and minimum phase. Furthermore, it is easy to see that the truncated model is a 'bad' approximation of the true model, unless (A_{12}, B_2) and/or (A_{21}, C_2) are 'small'.

Since the system matrices are unique only up to basis transformations (4), it is important to select a suitable realisation before truncation. In particular, balanced realisations have been proposed by many authors, because of their good behaviour from the point of view of model reduction. For an overview of balanced realisations see e.g. (Ober, 1996), (McGinnie, 1994).

Here the following two balancing schemes are considered:

Let the two Gramians (P, S) be defined as the solutions of the Lyapunov equations:

$$P = APA' + B\Sigma B' \tag{10}$$

$$S = \bar{A}' S \bar{A} + C'\Sigma^{-1}C \tag{11}$$

where $\bar{A} = A - BC$. The system (1) is said to be in *minimum phase balanced form* iff

$$P = S = \mathrm{diag}(\gamma_1, \ldots, \gamma_n) \tag{12}$$

where $\gamma_1 \geq \cdots \geq \gamma_n \geq 0$ are called the minimum phase singular values of the system. This scheme has been introduced and analysed in (McGinnie, 1994).

Consider the two Gramians $(P, \bar{P})$ defined by

$$P = APA' + B\Sigma B' \tag{13}$$
$$\bar{P} = A'\bar{P}A + \bar{C}'\bar{\Sigma}^{-1}\bar{C} \tag{14}$$

where $M = APC' + B\Sigma$, $\bar{C} = C - M'\bar{P}A$, $\bar{\Sigma} = \Sigma + CPC' - M'\bar{P}M$, and where $\bar{P}$ is the minimal solution of the above Riccati equation (14). The system (1) is said to be in *stochastically balanced form* iff

$$P = \bar{P} = \mathrm{diag}(\rho_1, \ldots, \rho_n) \tag{15}$$

where $1 \geq \rho_1 \geq \cdots \geq \rho_n \geq 0$ are the canonical correlation coefficients of the process (y_t). Note that $\bar{P}$ is the state variance of a (minimum phase) backward realisation of the process (y_t). This scheme has been proposed in (Desai and Pal, 1984).

There is a close relation between these two balancing schemes, as can be seen by the following lemma (a proof of this lemma for the continuous time case has been given in (McGinnie, 1994)):

Lemma 1. $S = (\bar{P}^{-1} - P)^{-1}$ and $\gamma_i^2 = \rho_i^2/(1-\rho_i^2)$.

This implies that minimum phase and stochastically balanced realisations are related to each other by a diagonal transformation T.

The next lemma collects some important properties of these two balancing schemes:

Lemma 2. (1) The γ_i's (ρ_i's) do not depend on the particular realisation of the system.
(2) The system is minimal iff $\gamma_n > 0$ ($\rho_n > 0$) holds.
(3) The balanced form is unique up to sign changes $T = \mathrm{diag}(\pm 1, \ldots, \pm 1)$, if all γ_i's (ρ_i's) are distinct.

Item (2) may be generalised in the sense, that the size of γ_{k+1} is a measure of the 'distance' of the transfer function $\mu(z)$ to the set of lower order transfer functions $\mathbb{M}(k)$. In particular, one can derive bounds for the approximation error of the truncated system, which depend on the singular values γ_j, $j > k$; see e.g. (McGinnie, 1994). However, these bounds refer to the continuous time case, and to the H_∞ norm.

Here the performance of these balanced truncation schemes will be evaluated with respect to the likelihood function (6). In particular, the case that the true transfer function is 'close' to the set of k-th order transfer functions $\mathbb{M}(k)$ is considered. To be more precise, a sequence of models converging to a system of order k will be considered. In order to make the exposition simpler, the analysis starts with a given sequence of realisations, rather than with a sequence of transfer functions.

Let $A(\varepsilon)$, $B(\varepsilon)$, $C(\varepsilon)$, $\Sigma(\varepsilon)$, be a sequence of realisations, which continuously depend on the scalar $\varepsilon \geq 0$.

Throughout the paper, it is assumed that:

$$\begin{array}{l} A(\varepsilon), \bar{A}(\varepsilon) \text{ are uniformly stable} \\ \Sigma(\varepsilon)^{-1} = O(1) \\ P_{11}^{-1}(\varepsilon) = O(1) \quad S_{11}^{-1}(\varepsilon) = O(1) \\ P_{22}(\varepsilon) = O(\varepsilon^2) \quad S_{22}(\varepsilon) = O(\varepsilon^2) \end{array} \tag{16}$$

Here and in the sequel, subscripts, like P_{11}, refer to a partitioning of the corresponding matrix conforming to (8). The notation $X(\varepsilon) = O(\varepsilon^s)$ means that there exist constants $\varepsilon_0 > 0$, and $c < \infty$ such that $\|X(\varepsilon)\| \leq c\varepsilon^s$ holds for all $\varepsilon_0 \geq \varepsilon \geq 0$. Furthermore $\|X\| = \sqrt{\lambda_{\max}(X'X)}$ denotes the 2-norm of the matrix X. Uniform stability is defined as follows: A (continuous) square matrix function $X(\varepsilon)$ is said to be uniformly stable iff there exist constants $\varepsilon_0 > 0$, $c < \infty$, and $\lambda < 1$ such that $\|X(\varepsilon)^k\| \leq c\lambda^k$ holds for all $k \geq 0$ and $\varepsilon_0 \geq \varepsilon \geq 0$.

Together these assumptions imply that

- the poles and zeros of the transfer functions are uniformly bounded away from the unit circle.
- the limiting transfer function is in $\mathbb{M}(k)$.
- the transfer functions are, in a certain sense, bounded away from the set of systems of order $s < k$.

The next lemma gives an equivalent formulation of the above assumptions.

Lemma 3. The assumptions (16) are equivalent to

$$\begin{array}{l} A_{11}(\varepsilon), A_{22}(\varepsilon), \bar{A}_{11}(\varepsilon) \text{ and} \bar{A}_{22}(\varepsilon) \\ \text{are uniformly stable} \\ \Sigma(0) > 0 \\ P_{11}(0) > 0 \qquad S_{11}(0) > 0 \\ A_{12}(\varepsilon) = O(\varepsilon) \quad A_{21}(\varepsilon) = O(\varepsilon) \\ B_2(\varepsilon) = O(\varepsilon) \quad C_2(\varepsilon) = O(\varepsilon) \end{array} \tag{17}$$

Note that the above realisations need not to be in balanced form. However, it is assumed that the lower right $(n-k) \times (n-k)$ blocks of the two gramians P and S converge to zero with the rate $O(\varepsilon^2)$. This will imply that the truncated model is a 'good' approximation of the true model. However, as will be shown in theorem 5, one gets better results if the gramians are 'almost' block diagonal. Such a block diagonal realisation may be obtained by the following lemma.

Lemma 4. Let a sequence of systems satisfying (16), and a sequence of (continuous) transformations satisfying

$$T = \begin{pmatrix} I & T_{12}(\varepsilon) \\ T_{21}(\varepsilon) & I \end{pmatrix} \quad \text{where } T_{12}(\varepsilon) = O(\varepsilon) \text{ and } T_{21}(\varepsilon) = O(\varepsilon). \tag{18}$$

be given.

(1) The transformed system defined by $\tilde{A} = TAT^{-1}$, $\tilde{B} = TB$, $\tilde{C} = CT^{-1}$ also satisfies (16).

(2) If the transformation is chosen to be $T_{12} = S_{11}^{-1}S_{12}$, $T_{21} = -P_{21}P_{11}^{-1}$, then

$$P_{12}(\varepsilon) = O(\varepsilon^3) \text{ and } S_{12}(\varepsilon) = O(\varepsilon^3) \quad (19)$$

holds.

Note that in general $P_{12} = O(\varepsilon)$ and $S_{12} = O(\varepsilon)$. If (19) holds, then the sequence of systems is said to be minimum phase block balanced.

In order to formulate the results, a parametrisation of the set $\mathbb{M}(k)$ of transfer functions of McMillan degree k is needed.

Let π denote the $k(k+2p)$ dimensional vector of stacked entries of the matrices (F,G,H). There is a mapping $\bar{\phi} : \pi \mapsto \theta = \bar{\phi}(\pi) \in \mathbb{R}^{2kp}$, which attaches a vector of parameters to any realisation. Of course $\bar{\phi}(\pi_1) = \bar{\phi}(\pi_2)$ holds iff the transfer functions corresponding to π_1 and π_2 are identical. Conversely, there is a mapping $\phi : \theta \mapsto \pi = \phi(\theta)$, which attaches a particular realisation to a vector of parameters. E.g. one could use a parametrisation based on balanced forms, see e.g. (Ober, 1996). These mappings are compatibel in the sense $\phi(\bar{\phi}(\theta)) = \theta$.

In particular let $\theta_0 = \bar{\phi}(\pi_0)$, where π_0 corresponds to the limit $(F(0),G(0),H(0)) = (A_{11}(0),B_1(0),C_1(0))$. Then it sufficient that the above mappings are defined and smooth in an open neighbourhood of π_0 and θ_0.

Note that the likelihood function depends on the parameters of the transfer function π (θ), on the variance Ω, as well as on the data generating process (A,B,C,Σ). Therefore, let Π denote the stacked vector of entries in the matrices (A,B,C,Σ). Furthermore let Π_0 correspond to the limit $(A(0),\ldots,\Sigma(0))$.

In order to simplify the expositions, only the simplified model reduction problem with fixed $\Omega = \Sigma$ is considered. This simplification is justified by Theorem 5, item (1). Therefore, with a slight abuse of notation, let

$$L(\theta;\Pi) = l(\phi(\theta);\Pi) = l(F,G,H,\Sigma;A,B,C,\Sigma) \quad (20)$$

Let $\bar{\theta}$ and $\hat{\theta}$ denote the parameters of the truncated and of the optimal system respectively. In order to assess the distance $(\bar{\theta} - \hat{\theta})$, the following regularity assumptions will be imposed on L:

There exist an open neighbourhood $(\mathcal{U} \times \mathcal{V}) \subseteq (\mathbb{R}^{2kp} \times \mathbb{R}^{(n+p)^2})$ of (θ_0,Π_0) such that

- the likelihood function has a unique global minimum for all $\Pi \in \mathcal{V}$
- the inverse of the Hessian of L is bounded for all $(\theta,\Pi) \in (\mathcal{U} \times \mathcal{V})$.

Now the main result of the paper is as follows:

Theorem 5. Consider a sequence of systems satisfying assumptions (16), and assume that the above regularity conditions on the likelihood function are fulfilled. Furthermore, let $s = 0$ in general, and $s = 2$ if the systems are minimum phase block balanced.

(1) $\mathrm{E}\, v_t v_t' = \Sigma + O(\varepsilon^4)$.
(2) The gradient of the likelihood function satisfies: $\frac{\partial L}{\partial \theta}(\bar{\theta},\Pi) = O(\varepsilon^{2+s})$
(3) $(\bar{\theta} - \hat{\theta}) = O(\varepsilon^{2+s})$.

The first item shows that the truncated model is a 'good' approximation of the true model, provided that the lower right blocks of the gramians are 'small'. This can be only achieved, when the $(n-k)$ smallest singular values γ_j, $j \geq k$ are 'small', and when a suitable realisation has been chosen. On the other hand, it follows from items (2) and (3) that, by picking a minimum phase balanced realisation, the truncated system will converge very fast to the optimal reduced order system. In this sense, minimum phase balanced truncation is locally optimal. Note that the truncated systems corresponding to a block balanced realisation and to a balanced realisation respectively, are related to each other by a state space transformation and thus represent the same k-th order transfer function. Hence, for the above results only an approximate block balanced realisation is needed.

By Lemma 1, it is clear that the above Theorem and thus the same local optimality property hold true for stochastically balanced truncation.

3. EXAMPLE

In this section the results obtained will be illustrated by the simple second order SISO ($p = 1$) systems:

$$\left.\begin{array}{c|c} A(\varepsilon) & B(\varepsilon) \\ \hline C(\varepsilon) & 1 \end{array}\right| = \left.\begin{array}{cc|c} -0.25 & 0.7\varepsilon & 2.3 \\ -0.7\varepsilon & 0.5 & 0.3\varepsilon \\ \hline -0.15 & -0.05\varepsilon & 1 \end{array}\right|$$

with $\Sigma = 1$. The goal is to find an optimal first order system ($k = 1$). We will compare the performance of the truncated system $\bar{\theta} = (\bar{f},\bar{g},\bar{h}) = (-0.25, 2.3, -0.15)$, with the minimum phase balanced truncated system $\tilde{\theta} = (\tilde{f},\tilde{g},\tilde{h})$, and the optimal first order model $\hat{\theta} = (\hat{f},\hat{g},\hat{h})$. Notice that the latter two reduced models depend on ε, whereas $\bar{\theta}$ does not depend on ε, and is equal to the limit θ_0. In this simple example, one may use $\theta = (f, f-gh)$, i.e. the pole and the zero of the transfer function, as parameters.

Figure 1 shows the partial derivative of the likelihood function with respect to f, for the two truncated models. Figure 2 shows the distance of the poles $\bar{f}$, $\tilde{f}$ of the two truncated model to the pole of the optimal model $\hat{f}$. In both plots a log-log scale is used, such that the rate of convergence can be easily seen. Note that one obtains similar pictures for the other partial derivatives and for the zero of the reduced order system.

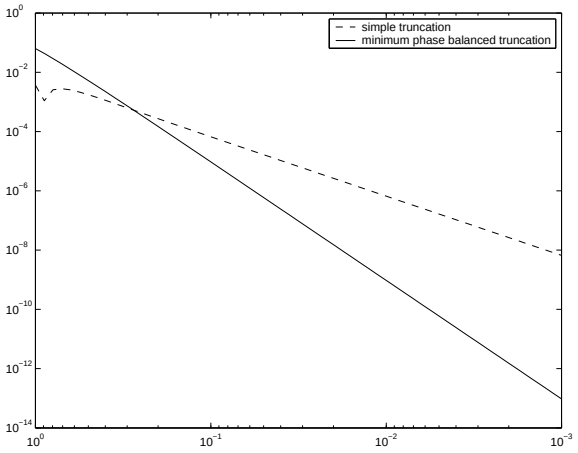

Fig. 1. Absolute value of the partial derivative of the likelihood function with respect to f, as a function of ε. The dashed line corresponds to the truncated model, and the solid line to the minimum phase balanced truncated model.

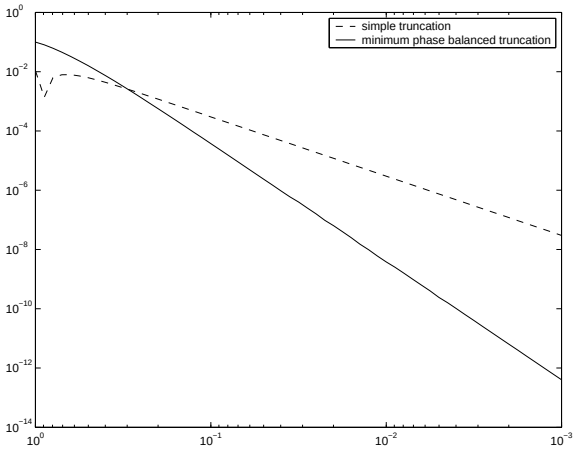

Fig. 2. Absolute difference of the poles of the truncated model (dashed line), and of the minimum phase balanced truncated model (solid line) with respect to the pole of the optimal reduced order model.

4. CONCLUSIONS AND REMARKS

In this paper the problem of finding a simple state space model for an ARMA process is considered, where the approximation error is measured by the asymptotic Gaussian likelihood. It has been shown that, given certain regularity conditions, the model obtained by minimum phase balanced truncation (and equivalently by stochastically balanced truncation) converges to the optimal reduced order model with a rate $O(\varepsilon^4)$, when the $(n-k)$ smallest minimum phase singular values of the true n-th order system converge to zero with rate $O(\varepsilon^2)$. For other truncation schemes, in general, only a rate $O(\varepsilon^2)$ will be attained. In this sense, minimum phase balanced truncation is 'locally optimal'. Sloppy speaking this means, minimum phase balanced truncation is 'almost' optimal if the true n-th order system is 'close' to the set of k-th order systems.

This result has a close relation to the so called CCA subspace estimation method. It has been shown in (Dahlen and Scherrer, 2001) that the CCA subspace is asymptotically equivalent to the following two step procedure (see also (Dahlen, 2001)):

(1) a (long) autoregressive model is estimated from the data.
(2) a stochastically balanced truncation of this AR model gives the desired estimate of the state space model generating the data.

Therefore, in view of the results obtained here, the second step, in a certain sense, is close to optimal with respect to the likelihood function.

5. PROOFS

PROOF of Lemma 1. Let

$$\begin{aligned} \mathcal{O}' &= (C', A'C', \ldots) \\ Y_+' &= (y_t', y_{t+1}', \ldots) \\ \Gamma_+ &= \mathrm{E}\, Y_+ Y_+' \end{aligned}$$

and let the infinite block toeplitz matrix $\mathcal{E}$ be defined by its block entries: $\mathcal{E}_{ij} = 0$, for $j > i$, $\mathcal{E}_{ii} = I$ and $\mathcal{E}_{ij} = CA^{i-j-1}B$ for $i > j$. Furthermore, an infinite block diagonal matrix $\mathcal{S}$ is defined by $\mathcal{S}_{ii} = \Sigma$. Then it is straightforward to see that

$$\Gamma_+ = \mathcal{O} P \mathcal{O}' + \mathcal{E}\mathcal{S}\mathcal{E}'$$

The matrix inversion lemma then gives

$$\begin{aligned} \mathcal{O}'(\mathcal{E}\mathcal{S}\mathcal{E}')^{-1}\mathcal{O} = \mathcal{O}'\Gamma_+^{-1}\mathcal{O} - \\ \mathcal{O}'\Gamma_+^{-1}\mathcal{O}(P^{-1} - \mathcal{O}'\Gamma_+^{-1}\mathcal{O})^{-1}\mathcal{O}'\Gamma_+^{-1}\mathcal{O}. \end{aligned}$$

In (Lindquist and Picci, 1996) it is proved that $\mathcal{O}'\Gamma_+^{-1}\mathcal{O} = \bar{P}$, and therefore, again by the matrix inversion lemma, one obtains

$$\mathcal{O}'(\mathcal{E}\mathcal{S}\mathcal{E}')^{-1}\mathcal{O} = (\bar{P}^{-1} - P)^{-1}$$

Finally it is easy to prove that $\mathcal{E}^{-1}\mathcal{O} = (C', \bar{A}'C', \ldots)'$ and thus

$$\mathcal{O}'(\mathcal{E}\mathcal{S}\mathcal{E}')^{-1}\mathcal{O} = \sum_{j\geq 0} (\bar{A}')^j C'\Sigma^{-1}C\bar{A}^j = S.$$

□

The main tool in proving Theorem 5 is the following lemma:

Lemma 6. Let A, F be two square matrices satisfying $||A^k|| \leq c\lambda^k$, $||F^k|| \leq c\lambda^k$, for some $\lambda < 1$, then the solution of the generalised Lyapunov equation

$$X = AXF + Q$$

satisfies $||X|| \leq (c^2/(1-\lambda^2))||Q||$.

PROOF.

$$\|X\| = \|\sum_{k\geq 0} A^k Q F^k\| \leq \sum_{k\geq 0} c^2 \lambda^{2k} \|Q\|$$

PROOF of Theorem 5. First a general expression for the gradient of the likelihood function is derived. Let $\bar{F} = F - GH$, $\bar{L}(z) = (z^{-1}I - \bar{F})^{-1}$ and $w_t = \bar{L}(z)Gy_t$, then the total derivative of the prediction error v_t with respect to the entries in F, G, and H is given by:

$$\begin{aligned} v_t &= \nu^{-1}(z)y_t = (I - H\bar{L}(z)G)y_t \\ -\mathrm{d}v_t &= H\bar{L}(z)(\mathrm{d}F - G\,\mathrm{d}H)w_t \\ &\quad + H\bar{L}(z)\,\mathrm{d}Gv_t + \mathrm{d}Hw_t \end{aligned} \tag{21}$$

A state space model for $(v_t', w_t')'$ (with inputs (u_t)) is given by:

$$\left.\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right. := \left.\begin{array}{cc|c} A & 0 & B \\ GC & \bar{F} & G \\ \hline 0 & I & 0 \\ C & -H & I \end{array}\right. \tag{22}$$

Given this model, define:

$$\begin{aligned} \hat{P} &= \hat{A}\hat{P}\hat{A}' + \hat{B}\Sigma\hat{B}' \\ \hat{M} &= \hat{A}\hat{P}\hat{C}' + \hat{B}\Sigma\hat{D}' \\ X &= \bar{F}'X\hat{A} + H'\Omega^{-1}(C, -H) \\ Y &= X\hat{M} = (Y_1, Y_2), \qquad Y_1 \in \mathbb{R}^{k\times k} \\ \Gamma_{vv} &= \mathrm{E}\,v_t v_t' = (C, -H)\hat{P}(C, -H)' + \Sigma \\ \Gamma_{vw} &= \mathrm{E}\,v_t w_t' = (C, -H)\hat{P}(0, I)' \end{aligned}$$

Then, by standard calculus, the following expression for the total derivative of the likelihood may be obtained:

$$\begin{aligned} -(1/2)\,\mathrm{d}l = -\mathrm{E}\,v_t'\Omega^{-1}\,\mathrm{d}v_t = \\ \mathrm{E}(v_t'\Omega^{-1}\sum_{j>0} L_j'\,\mathrm{d}z_{t-j} + v_t'\Omega^{-1}\,\mathrm{d}Hw_t) = \\ \mathrm{tr}(Y'(\,\mathrm{d}F - G\,\mathrm{d}H, \mathrm{d}G)) + \mathrm{tr}(\Gamma_{wv}\Omega^{-1}\,\mathrm{d}H) \end{aligned} \tag{23}$$

Here $L_j' = H\bar{F}^{j-1}$, $\mathrm{d}z_t = (\,\mathrm{d}F - G\,\mathrm{d}H)w_t + \mathrm{d}Gv_t$ and

$$\begin{aligned} \sum_{j>0} L_j\Omega^{-1}\,\mathrm{E}\,v_{t+j}(w_t', v_t')' = \\ \sum_{j\geq 0} (\bar{F}')^j H'\Omega^{-1}(C, -H)\hat{A}^j\hat{M} = \\ X\hat{M} = Y \end{aligned}$$

have been used.

Now consider the case of a sequence of data generation models (16), and where the reduced order model is obtained by truncation, and where $\Omega = \Sigma$. In this case the model (22) (after a suitable state transformation) may be written as:

$$\left.\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right. := \left.\begin{array}{ccc|c} A_{11} & A_{12} & 0 & B_1 \\ A_{21} & A_{22} & \bar{0} & B_2 \\ 0 & -A_{12} & \bar{A}_{11} & 0 \\ \hline I & 0 & I & 0 \\ 0 & C_2 & -C_1 & I \end{array}\right. \tag{24}$$

By Lemma 6, one obtains

$$\hat{P} = \begin{pmatrix} P_{11} & O(\varepsilon^{1+s}) & O(\varepsilon^{2+s}) \\ O(\varepsilon^{1+s}) & O(\varepsilon^2) & O(\varepsilon^3) \\ O(\varepsilon^{2+s}) & O(\varepsilon^3) & O(\varepsilon^4) \end{pmatrix}$$

$$\hat{M} = \begin{pmatrix} A_{11}P_{11} + O(\varepsilon^{2+s}) & B_1\Sigma + O(\varepsilon^{2+s}) \\ A_{21}P_{11} + O(\varepsilon^{1+s}) & B_2\Sigma + O(\varepsilon^3) \\ O(\varepsilon^{2+s}) & O(\varepsilon^4) \end{pmatrix}$$

$$X = (O(\varepsilon^{2+s}), O(\varepsilon^{1+s}), -S_{11} + O(\varepsilon^{2+s})).$$

Hence, it follows that $\Gamma_{vv} = \Sigma + O(\varepsilon^4)$, $\Gamma_{vw} = O(\varepsilon^{2+s})$ and $Y = O(\varepsilon^{2+s})$. This proves items (1), and (2), by using (23) and the following relation

$$\frac{\partial L}{\partial\theta}(\bar{\theta}, \Pi) = \frac{\partial l}{\partial\pi}\frac{\partial\bar{\phi}}{\partial\theta}.$$

By the mean value theorem, it follows that

$$\frac{\partial L}{\partial\theta}(\bar{\theta}, \Pi)(\bar{\theta} - \hat{\theta}) = (\bar{\theta} - \hat{\theta})'\frac{\partial^2 L}{\partial\theta\partial\theta'}(\theta^*, \Pi)(\bar{\theta} - \hat{\theta})$$

where $\theta^* = \alpha\bar{\theta} + (1-\alpha)\hat{\theta}$, for some $0 \leq \alpha \leq 1$. Now the boundedness of the inverse of the Hessian implies item (3). □

6. REFERENCES

Dahlen, A. (2001). Identification of stochastic systems: Subspace methods and covariance extension. PhD thesis. Royal Institute of Technology. Stockholm, Sweden.

Dahlen, A. and W. Scherrer (2001). The relation of the cca subspace method to a balanced reduction of an autoregressive model. submitted to Journal of Econometrics.

Desai, U. B. and D. Pal (1984). A transformation approach to stochastic model reduction. *IEEE Transactions on Automatic Control* **AC-29**(12), 1097–1100.

Lindquist, A. and G. Picci (1996). Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica* **32**(5), 709–733.

McGinnie, B.P. (1994). A Balanced View of System Identification. PhD thesis. Cambridge, University.

Ober, R. J. (1996). Balanced canonical forms. In: *Identification, Adaptation, Learning, The Science of Learning Models from Data* (S. Bittanti and G. Picci, Eds.). Vol. 153 of *NATO ASI Series F: Computer and System Sciences*. pp. 120–183. Springer-Verlag. Berlin.