APPROXIMATIONS FOR THE BIAS AND THE VARIANCE OF A CLASS OF NONRECURSIVE AUTOREGRESSIVE ESTIMATORS

D. PROSS

Institut für Prozeßmeßtechnik und Prozeßleittechnik,
Universität Karlsruhe,
German Federal Republic

Summary

The quality of autoregressive estimators can be judged by statistical criteria like the bias and the variance of the estimated parameters. But generally, analytical expressions for these criteria exist only in the asymptotic case of great samples. — This paper presents a class of nonrecursive autoregressive estimators which contains most of the usually used autoregressive estimators, especially the Orthogonal Regression. A statistical linearization of this class leads to approximations for the quality criteria. They are valid for every sample size, but only for small disturbances. Further, there are hints about a good estimation depending on the effective sampling time in the autoregressive equation.

Introduction and Overview

Many parameter estimation methods for autoregressive models are in use which produce satisfactory results (e.g. Graupe et al., 1980; Haykin, 1979). The following reasons are noteworthy for this contribution.

If we will prove or compare the quality of the various methods for example by means of the bias and the variance of the estimated parameters, in general, we find no analytical expressions for them. However, the asymptotic behaviour is known for great samples. For medium and small samples we have to do long and tedious test series.

Further, the criteria show interesting dependences from free parameters, e.g. the effective sampling time in the autoregressive equation. It would be of great advantage to use this dependence for improving the estimators. But by trial and error the variation of the parameters is a matter of experience and of good luck.

Therefore, analytical expressions for the quality criteria are desirable. They can give insight into the proceedings and characteristics of the estimators. Now it would be possible to improve the estimators by a systematic choice of the free parameters.

The starting point of our approach is the description of a measured signal by a special exogenous model permitting both stationary and nonstationary processes. After the introduction of discrete time the autoregressive model will be derived from the exogenous one by a linear transformation and a constraint. The autoregressive model obtained by this way can be compared with the solution of a homogeneous linear difference equation having constant coefficients.

These coefficients may be estimated by many methods. From a general description of the estimation problem it results in a class of nonlinear nonrecursive autoregressive estimators. It comprises most of the usually used methods, especially the Orthogonal Regression.

The bias and the variance are employed as quality criteria for the estimated parameters. We have to deal with the expectation operator, therefore linearization seems to be the only reasonable way to master the nonlinear problem. It appears as a small signal approximation relative to the noise for the quality criteria. We obtain simple expressions in the case of Gaussian white noise. For further simplification the covariance matrix of the estimated parameter vector needs the additional assumption of noncorrelated measurement values in the measurement matrix $\mathbf{Y}$. In the last section, the validity of the approximations is shown by an example which is compared to the statistics of a computer-simulated test series.

Parametric Models for Signals

Basically, the problem is assumed to have a series of samples and the task is to analyse the signal which caused the record.

The exogenous model

As a first step we fit for the continuous time t a mathematical parametric model $\mathring{y}(t)$ on the output signal $y(t)$ of a process. The process may be disturbed by an additive measurement noise $\Delta y(t)$. This arrangement is termed a feed forward or output error model (Fig. 1).

It can be written

$$y(t) = \mathring{y}(t) + e(t) \tag{2.1}$$

wherein $e(t)$ is the equation error and contains both model and measurement errors. However, only errors due to measurement are taken into consideration, because the model error cannot be treated in a unified and systematic way.

As a special exogenous model we use

$$\mathring{y}(t) = \sum_{j=1}^{p} r_j \cdot \exp(s_j t). \tag{2.2}$$

A general exogenous model represents a signal by a weighted sum of functions or elements of a function system. The orthogonal system of the complex exponential functions is chosen, because of the advantages:

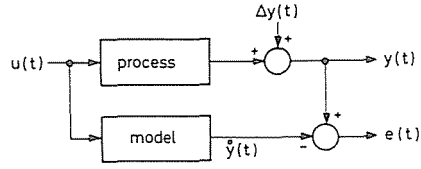

Fig. 1. Feed forward or output error model

a) the physical meaning of its parameters:

r_j as a complex amplitude;
s_j as a complex frequency.

b) it can be interpreted as the solution of an inhomogeneous differential equation. This equation describes a linear, time invariant or quasistationary system with lumped parameters excited by an input signal. Because we do not allow model errors and we have a linear system the input signal must be composed like Eq. (2.2). Quasistationary means that the parameters r_j and s_j should be constant during the time of one measurement series.

Interpreting the exogenous model as a solution of an inhomogeneous differential equation having a transient term and a steady-state term, the model can be conveniently used for both stationary and nonstationary signals.

The structural parameter p which is termed the model signal order is assumed to be a known parameter. The frequencies s_j contain the characteristic frequencies of the signal composed by the characteristic frequencies of the system, as well as the input signal. For their appearing in the output signal it is necessary that the system frequencies are excited by the input signal and the input frequencies are transferred by the system. We assume that all complex frequencies s_j are simple and different for identification.

Working with the computer we consider the discrete time and sample the signal $y(t)$ at a distance Δ

$$y(t = i\Delta) = y(i) = \sum_{j=1}^{p} r_j \exp(s_j i\Delta) + e(i) \tag{2.3}$$

for $i = 0, 1, \ldots, N_M - 1$, where N_M is the number of the samples.

In the comfortable matrix representation we get for all sampling instants

$$\mathbf{y}=\mathbf{X}\mathbf{b}+\mathbf{e} \tag{2.4}$$

with the following components:

$$(\mathbf{y})_i=y(i) \tag{2.5a}$$

$$(\mathbf{X})_{ij}=\exp(s_j i\Delta)=q_j^i \tag{2.5b}$$

$$(\mathbf{b})_j=r_j \tag{2.5c}$$

$$(\mathbf{e})_i=e(i) \tag{2.5d}$$

$$\text{for } i=0, 1, \ldots, N_M-1 \text{ and } j=1, 2, \ldots p.$$

If all s_j are different with respect to each other, the matrix $\mathbf{X}$ possesses the full rank p, its columns span a p-dim. vector space. — As a vector space we suppose a Hilbert space with complex or real elements.

For analyzing the signal $y(i)$ we have the problem of determining the coupled set of the $2p$ parameters: r_j and s_j. The normally used least-squares criterion produces $2p$ nonlinear and transcendental equations which can not be solved analytically.

The autoregressive model

To overcome the problem due to the coupled parameters we apply a linear transformation subject to a constraint on our special exogenous model and find

$$\mathbf{A}\mathbf{y}=\mathbf{A}\mathbf{e}=\boldsymbol{\varepsilon} \tag{2.6}$$

with the constraint

$$\mathbf{A}\mathbf{X}=\mathbf{0}. \tag{2.7}$$

By this way the parameters r_j are eliminated, the constraint decouples the parameter set. The constraint plays the important role of a connection between the autoregressive and the exogenous signal representation. Every row in the equation system (2.6) shows a stochastic difference equation or an autoregressive equation of unknown order for the erroneous measured values $y(i)$.

If we agree upon an equal distance Δ_{eff} between the values $y(i)$ in the equations we can write with δ as a sampling factor and Δ as the original sampling time

$$\Delta_{\text{eff}}=\delta\Delta. \tag{2.8}$$

Therefore, any row of the system (2.6) represents the identical difference equation but for other, say another following time instants.

To determine the necessary order of the autoregressive model we examine one entry of the matrix constraint. It results in a polynomial with the coefficients $a_k(\delta)$ and p different roots

$$q_j^{\delta} = \exp(s_j \Delta\delta). \tag{2.9}$$

The polynomial is independent of time and because of the p roots we choose the order as to be p:

$$\sum_{k=0}^{p} a_k(\delta) q_j^{\delta \cdot k} = 0, \tag{2.10}$$

which is the characteristic equation. Further, we find a difference equation of order p

$$\sum_{k=0}^{p} a_k(\delta) y(i + k\delta) = \varepsilon(i) \tag{2.11}$$

for $i = 0, 1, \ldots N-1$. N means the number of the possible difference equations and is given by

$$N = N_M - p\delta. \tag{2.12}$$

In the undisturbed case we have a homogeneous difference equation for the values $\mathring{y}(i + k\delta)$.

From another point of view we can interpret our proceeding up to here as tracing back to the solution of a difference equation. Our starting point has been the solution itself or the exogenous model. Its terms are caused by the exponential solution set-up and the roots of the resulting characteristic equation. The characteristic equation corresponds to the constraint and because we wish a homogeneous difference equation — possessing only information about the output values — we choose a homogeneous constraint.

It is naturally possible to derive the autoregressive model by means of system theory, especially the Z-transform (Stöckle, 1984).

The demonstrated way is preferred because of the relative simple connection of the parameters in our case, and of the explicit dependence from the effective sampling time.

In summary, we can identify the exogenous model by the following steps:

a) From the autoregressive equation system (2.6) we compute the coefficients $a_k(\delta)$.
b) We set up a polynomial with these coefficients (2.10), the roots of it (2.9) imply the frequencies s_j.
c) In the last step we determine the weighting coefficients r_j e.g. by the well-known methods of the regression theory.

With a little knowledge about the input signal, for example its shape, or about the system, it is often possible to divide the identified frequencies s_i of the output signal in a part belonging to the input signal and a part belonging to the

system. In this case we can attain a complete parameter identification of the input signal as well as of the system.

Now we direct our attention to the computation of the coefficients of the autoregressive equation.

A Class of Nonrecursive Autoregressive Estimators

The general autoregressive estimator

Because the measurement values $y(i)$ are erroneous we need an overdetermined equation system for a good estimation of the parameters $a_k(\delta)$. If we collect them in a vector $\mathbf{a}$ and the values $y(i)$ in a matrix $\mathbf{Y}$ with

$$(\mathbf{Y})_{ik} = y(i + k\delta), \tag{3.1}$$

the equation (2.6) can be modified as

$$\mathbf{Y}\mathbf{a} = \boldsymbol{\varepsilon}. \tag{3.2}$$

Using the least-squares criterion we have to minimize the normalized error function

$$F(\mathbf{a}) = \frac{1}{N}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \mathbf{a}^T\frac{\mathbf{Y}^T\mathbf{Y}}{N}\mathbf{a} = \mathbf{a}^T\mathbf{M}\mathbf{a}. \tag{3.3}$$

The matrix $\mathbf{M}$ can be called an empirical correlation matrix because its elements are defined as

$$(\mathbf{M})_{kl} = \frac{1}{N}\sum_{i=0}^{N-1} y(i + k\delta)\cdot y(i + l\delta). \tag{3.4}$$

The error function is homogeneous and we receive the minimum at the trivial solution $\hat{\mathbf{a}} = 0$. Preventing that we have to introduce a normalization for $\mathbf{a}$, which may lead to many solutions but not all of them are good.

A general normalization (Kronmüller, 1978) can be formulated as a linear constraint on $\mathbf{a}$

$$\mathbf{a}^T\mathbf{g} = 1, \quad |\mathbf{g}| = 1, \tag{3.5}$$

where $\mathbf{g}$ is provisionally a general weighting vector for the coefficients a_k, and assumed to have a suitable length of unity.

This statement comprises most of the generally used normalizations. They can be treated all at once in this way, the symmetry of the minimum problem (3.3) is preserved and it is possible to find an optimal weighting vector.

We have to minimize now an expanded error function which pays regard to the constraint

$$\mathrm{LG}(\mathbf{a}, \lambda) = F(\mathbf{a}) - 2\lambda(\mathbf{a}^T\mathbf{g} - 1) = \mathbf{a}^T\mathbf{M}\mathbf{a} - 2\lambda(\mathbf{a}^T\mathbf{g} - 1) \tag{3.6}$$

and we find

$$\mathbf{M}\hat{\mathbf{a}} = \hat{\lambda}\mathbf{g}. \tag{3.7}$$

Using the constraint (3.5) we get

$$\hat{\lambda} = \frac{1}{\mathbf{g}^T\mathbf{M}^{-1}\mathbf{g}} = \hat{\mathbf{a}}^T\mathbf{M}\hat{\mathbf{a}} = F(\hat{\mathbf{a}}). \tag{3.8}$$

We see that the Lagrangian multiplier $\hat{\lambda}$ has a clear meaning, it is the minimized error sum which depends on the choice of the weighting vector.

From equation (3.8) we obtain

$$\hat{\mathbf{a}} = \hat{\lambda}\mathbf{M}^{-1}\mathbf{g} = \frac{1}{\mathbf{g}^T\mathbf{M}^{-1}\mathbf{g}}\mathbf{M}^{-1}\mathbf{g} = \frac{\mathbf{M}_{ad}\mathbf{g}}{\mathbf{g}^T\mathbf{M}_{ad}\mathbf{g}}, \tag{3.9}$$

the last reformulation with the adjugate matrix $\mathbf{M}_{ad}$ resulting from

$$\mathbf{M}^{-1} = \frac{1}{\det \mathbf{M}}\mathbf{M}_{ad} \tag{3.10}$$

is useful if the matrix $\mathbf{M}$ becomes singular.

The optimum weighting vector $\mathbf{g}$: *the Orthogonal Regression*

The formula (3.8) for the minimized error function shows a dependence from the weighting vector. We find the absolute minimum of the criterion with the help of the Rayleigh-coefficient well known in linear algebra (Strang, 1980). The optimum weighing vector $\mathbf{g}_{\mathrm{opt}}$ has to be

$$\mathbf{g}_{\mathrm{opt}} = \mathbf{l}_0 \tag{3.11}$$

that means the normalized characteristic vector belonging to the smallest characteristic value of the matrix $\mathbf{M}$. If we substitute this optimum vector into the estimator (3.9) it turns out that

$$\hat{\mathbf{a}}_{\mathrm{opt}} = \mathbf{l}_0. \tag{3.12}$$

Therefore the constraint preventing the trivial solution reduces to a simple normalization rule.

This optimum estimation method is known as Orthogonal Regression or Eigenvector method. It can be found in the literature e.g. (Koopmans, 1936; Levin, 1964; Kronmüller, 1978).

Some geometrical considerations

For enlightening our subject from another point of view, we start some geometrical considerations.
In the undisturbed case the system of equations (3.2) becomes

$$\mathring{\mathbf{Y}}\mathbf{a}=0 \tag{3.13}$$

and contains N error-free difference equations according to (2.11):

$$\sum_{k=0}^{p} a_k(\delta)\cdot \mathring{y}(k+i)=0=\mathbf{a}^T\mathring{\mathbf{y}}_Z(i). \tag{3.14}$$

In the $(p+1)$-dim. Hilbert space every row vector $\mathring{\mathbf{y}}_Z(i)$ of the matrix $\mathring{\mathbf{Y}}$ is orthogonal to the coefficient vector $\mathbf{a}$. The row vectors span a p-dim. subspace $R(\mathring{\mathbf{Y}})$ – the row space of the matrix $\mathring{\mathbf{Y}}$ – and the vector $\mathbf{a}$ is the associated normal vector in this case (see Fig. 2) which is termed as the exact coefficient vector $\mathring{\mathbf{a}}$.

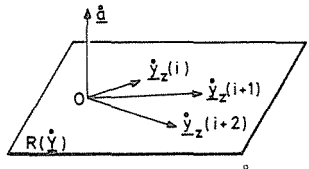

Fig. 2. The subspace $R(\mathring{\mathbf{Y}})$ and its normal vector $\mathring{\mathbf{a}}$

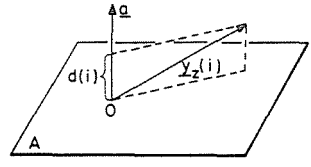

Fig. 3. The distance $d(i)$ of the vector $\mathbf{y}_z(i)$ from the hyperplane A

The rule (3.3) for computing the matrix $\mathring{\mathbf{M}}=\frac{1}{N}\mathring{\mathbf{Y}}^T\mathring{\mathbf{Y}}$ indicates that the rows and columns of $\mathring{\mathbf{M}}$ are linear combinations of the rows of $\mathring{\mathbf{Y}}$, therefore the row- and the column-space of $\mathring{\mathbf{M}}$ are identical with $R(\mathring{\mathbf{Y}})$.

In the case of disturbed values $y(i)$ we deal with the equation system (3.2) which contains N inner products. We can express them as

$$\mathbf{a}^T\mathbf{y}_Z(i)=|\mathbf{a}|\,|\mathbf{y}_Z(i)|\cos(\mathbf{a},\mathbf{y}_Z(i))=|\mathbf{a}|\,d(i)=\varepsilon(i), \tag{3.15}$$

where $d(i)$ can be interpreted as the distance of the tip of the vector $\mathbf{y}_Z(i)$ from a hyperplane A characterized by the normal vector $\mathbf{a}$ (Fig. 3).

Considering the constraint (3.5) we get for the length of $\mathbf{a}$ a term (Fig. 4) which is useful for the equation error

$$\varepsilon(i)=\frac{d(i)}{|\cos(\mathbf{a},\mathbf{g})|}. \tag{3.16}$$

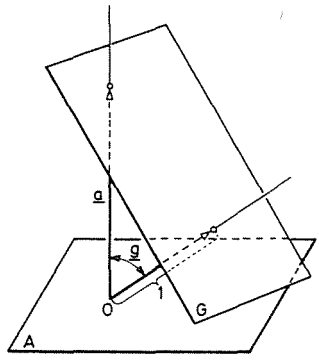

Fig. 4. The length of the coefficient vector **a** constrained by $\mathbf{a}^T\mathbf{g}=1$

The results yield that every equation error depends on the angle between the coefficient vector **a** and the weighting vector **g**. If the vectors are parallel the absolute minimum is reached. In this case the direction of **a** is derived by minimizing the sum of the squared *orthogonal* distances $d(i)$ of the vectors $\mathbf{y}_z(i)$ from the hyperplane A or regression plane: "Orthogonal Regression".

Quality Criteria for the Estimated Parameter Vector â

The definition of the bias and the variance

The derivation of the class of nonrecursive autoregressive estimators has taken place without any statistical means. As quality criteria we choose the bias and the variance usually used in statistics. With $E\{\ldots\}$ as expectation operator and splitting up the estimated vector â into the exact part å and an error part $\Delta\mathbf{a}$

$$\hat{\mathbf{a}}=\mathring{\mathbf{a}}+\Delta\mathbf{a} \tag{4.1}$$

we can define the bias

$$E\{\Delta\mathbf{a}\}=E\{\hat{\mathbf{a}}-\mathring{\mathbf{a}}\} \tag{4.2}$$

and the covariance matrix of the estimation error

$$\mathbf{V}_{\hat{\mathbf{a}}}=E\{(\hat{\mathbf{a}}-E\{\hat{\mathbf{a}}\})(\hat{\mathbf{a}}-E\{\hat{\mathbf{a}}\})^T\}. \tag{4.3}$$

The bias is the difference between the mean of the estimated and the exact vector, if we can repeat the estimation several times under the same circumstances.

The variance of the estimated parameters $\hat{a}_i$ is given by the diagonal entries of the matrix $\mathbf{V}_{\hat{a}}$, and they contain a hint about the scattering or dispersion of the estimated values about their means.

Because of the statistical criteria the measurement values $y(i)$ are treated as random variables. We assign the statistical properties of Gaussian white noise to the measurement errors $\Delta y(i)$:

$$\Delta y(i) \rightarrow N(0, \sigma^2). \tag{4.4}$$

The approximation for the bias of $\hat{\mathbf{a}}$

For the computation of the criteria we have to operate with the expectation on the vector $\hat{\mathbf{a}}$ and the product $\hat{\mathbf{a}}\hat{\mathbf{a}}^T$. Because the estimator (3.9) is nonlinear in a complicated way, we are forced to linearize. Every term is developed into a Taylor series about its exact value and the quadratic and higher order terms are neglected. Accordingly the small signal approximations are obtained as, e.g.

$$\mathbf{g} \approx \mathring{\mathbf{g}} + \Delta\mathbf{g} \tag{4.5}$$

$$\mathbf{M} \approx \mathring{\mathbf{M}} + \Delta\mathbf{M}. \tag{4.6}$$

Computations in linear algebra give the desired result, as,

$$\Delta\mathbf{a} \cong -\frac{1}{\mathring{\mathbf{g}}^T \mathring{\mathbf{l}}_0} \mathbf{P}\mathring{\mathbf{M}}^- \mathbf{P}\Delta\mathbf{M}\mathring{\mathbf{l}}_0, \tag{4.7}$$

where $\mathring{\mathbf{l}}_0$ denotes the normalized eigenvector belonging to the smallest eigenvalue of the matrix $\mathring{\mathbf{M}}$,

$$\mathbf{P} = \mathbf{I} - \frac{\mathring{\mathbf{l}}_0 \mathring{\mathbf{g}}^T}{\mathring{\mathbf{g}}^T \mathring{\mathbf{l}}_0} \text{ is the projection operator,} \tag{4.8}$$

$$\mathring{\mathbf{M}}^- = \sum_{i=1}^{p} \frac{1}{\mathring{\lambda}_i} \mathring{\mathbf{l}}_i \mathring{\mathbf{l}}_i^T \quad \text{is the singular value decomposition (Strang, 1980; Kuhnert, 1976).} \tag{4.9}$$

Also, $\mathring{\lambda}_i$ are the eigenvalues of $\mathring{\mathbf{M}}$ with $o < \mathring{\lambda}_1 < \mathring{\lambda}_2 < \ldots < \mathring{\lambda}_p$,

$\mathring{\mathbf{l}}_i$ are the according normalized eigenvectors.

In the case of the Orthogonal Regression with $\mathring{\mathbf{g}} = \mathring{\mathbf{l}}_0$ we attain

$$\Delta\mathbf{a}_{OR} = \Delta\mathbf{l}_0 \cong -\mathring{\mathbf{M}}^- \Delta\mathbf{M}\mathring{\mathbf{l}}_0. \tag{4.10}$$

Taking expectations produces the bias

$$E\{\Delta\mathbf{a}\} \cong \sigma^2 |\mathring{a}|^2 \mathbf{P}\mathring{\mathbf{M}}^{-}\mathring{\mathbf{g}}, \tag{4.11}$$

which vanishes for the Orthogonal Regression

$$E\{\Delta\mathbf{a}_{OR}\} = E\{\Delta\mathbf{l}_0\} \cong 0. \tag{4.12}$$

The results are illustrated in Fig. 5. In general the vectors $\mathring{\mathbf{a}}$ and $\hat{\mathbf{a}}$ touch the hyperplane G with their tips, the difference vector $\Delta\mathbf{a}$ lies entirely in the plane G

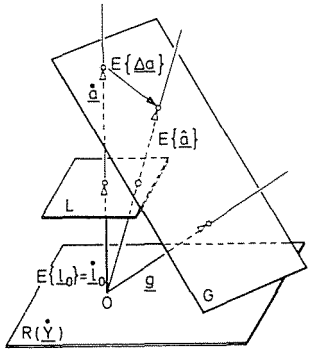

Fig. 5. The expectation of the linearized estimation errors $\Delta\mathbf{a}$ and $\Delta\mathbf{l}_0$

depending on the direction of the weighting vector $\mathbf{g}$. In the mean the variation of $\hat{\mathbf{a}}$ is not identical with $\mathring{\mathbf{a}}$, we have a bias $E\{\Delta\mathbf{a}\}$ which is always negative:

$$|E\{\hat{\mathbf{a}}\}| = |\mathring{\mathbf{a}}| \cdot \sqrt{1 - 2\sigma^2 |\mathring{\mathbf{a}}|^2 \mathring{\mathbf{g}}^T \mathring{\mathbf{M}}^{-} \mathring{\mathbf{g}}} \leq |\mathring{\mathbf{a}}|. \tag{4.13}$$

In the case of the Orthogonal Regression the estimated vector $\hat{\mathbf{a}}_{OR} = \mathbf{l}_0$ varies with its tip touching L, but in the mean there is no difference between the estimated and the exact vector.

The approximation for the covariance matrix of $\hat{\mathbf{a}}$

We can reformulate equation (4.3) for the covariance matrix in terms of the difference vector $\Delta\mathbf{a}$

$$\mathbf{V}_{\hat{\mathbf{a}}} = E\{(\Delta\mathbf{a} - E\{\Delta\mathbf{a}\})(\Delta\mathbf{a} - E\{\Delta\mathbf{a}\})^T\}. \tag{4.14}$$

If we put in $\Delta\mathbf{a}$ according to (4.7) we find

$$\mathbf{V}_{\hat{\mathbf{a}}} \cong |\mathring{\mathbf{a}}|^2 \mathbf{P}\mathring{\mathbf{M}}^{-}\mathbf{P}^T [E\{\Delta\mathbf{M}\mathring{\mathbf{l}}_0 \mathring{\mathbf{l}}_0^T \Delta\mathbf{M}\} - E\{\Delta\mathbf{M}\mathring{\mathbf{l}}_0\} E\{\mathring{\mathbf{l}}_0^T \Delta\mathbf{M}\}] \mathbf{P}\mathring{\mathbf{M}}^{-}\mathbf{P}^T. \tag{4.15}$$

Taking the expectation we attain complicated sums (Pross, 1984), thus in general only a numerical treatment is possible.

But if the limitation for the allowed region of the sampling factor δ related to the effective sampling time in the autoregressive equation is accepted the sums simplify drastically. The limitation means to have only noncorrelated measurement values in the matrix $\mathbf{Y}$:

$$N \leq \delta. \tag{4.16}$$

Thus we get,

$$\mathbf{V}_{\hat{\mathbf{a}}} \approx \frac{\sigma^2}{N} |\mathring{\mathbf{a}}|^2 \mathbf{P} \mathring{\mathbf{M}}^- \mathbf{P}^T \tag{4.17}$$

and in the special case of the Orthogonal Regression

$$\mathbf{V}_{\mathbf{l}_0} \approx \frac{\sigma^2}{N} \mathring{\mathbf{M}}^- = \frac{\sigma^2}{N} \cdot \sum_{i=1}^{p} \frac{1}{\mathring{\lambda}_i} \mathring{\mathbf{l}}_i \mathring{\mathbf{l}}_i^T. \tag{4.18}$$

The last formula is the lower boundary corresponding to the Cramér—Rao-inequality. The matrix $\mathring{\mathbf{M}}$ contains all information we have which is the Fisher information matrix (Sorenson, 1980). In this special case the estimation is equal to a maximum likelihood estimation, too (Koopmans, 1937).

The following remarks are worthy according to the covariance matrices,

a) they are singular, because they describe a singular or defective distribution. The components of the $(p+1)$-dim. random variable $\hat{\mathbf{a}}$ are linear dependent because of the normalization constraint.

b) It is difficult to compare matrices, especially singular ones. For that, we can define several criteria and one possibility is the average variance of all components $\hat{a}_i$. In spite of equal length of the normalized parameter vectors we find

$$\frac{\overline{\mathrm{var}(\hat{\mathbf{a}}_{\mathrm{norm}})}}{\overline{\mathrm{var}(\mathbf{l}_0)}} \approx \frac{1 + |\mathring{\mathbf{a}}|^2 \dfrac{\mathring{\mathbf{g}}^T \mathring{\mathbf{M}}^- \mathring{\mathbf{g}}}{tr \mathring{\mathbf{M}}^-}}{1 - 2\sigma^2 |\mathring{\mathbf{a}}|^2 \mathring{\mathbf{g}}^T \mathring{\mathbf{M}}^- \mathring{\mathbf{g}}} \geq 1, \tag{4.19}$$

the Orthogonal Regression has in the mean the smaller variance, $tr\mathring{\mathbf{M}}$ is the trace of the matrix $\mathring{\mathbf{M}}^-$.

c) The matrices $\mathbf{P}$ and $\mathring{\mathbf{M}}^-$ consist of normalized vectors, the amounts of their components are less or equal one. The amount of $|\mathring{\mathbf{a}}|$ can be controlled by the direction of the weighing vector $\mathbf{g}$, therefore only N and the second smallest eigenvalue $\mathring{\lambda}_1$ can have a large influence on the variance of $\hat{\mathbf{a}}$. We take N as a measure for the overdetermination of the equation system and $\mathring{\lambda}_1$ as a measure for the further degree of the linear dependence of the singular matrix $\mathring{\mathbf{M}}$. We have to consider that the smallest eigenvalue does not appear, it is zero in the

undisturbed case or very small representing only the noise of the measurement values and has been neglected in the singular value decomposition matrix.

d) The derived approximations contain only undisturbed and therefore unknown quantities. But applying ergodicity (Papoulis, 1965) these unknown quantities can be replaced in a good manner by the quantities which are computed from the disturbed measurement values.

An Example

As an example we deal with a first order system and a unit step as input signal. Therefore, we obtain an autoregressive equation of the order two

$$a_0(\delta) \cdot y(i) + a_1(\delta) \cdot y(i+\delta) + a_2(\delta) \cdot y(i+2\delta) = \varepsilon(\delta, i) \tag{5.1}$$

and estimate the coefficients by the Orthogonal Regression. Beginning with the step we take $N_M = 100$ samples during the measurement time T which equals the time constant T of the system. The samples are simulated with an additive white noise, its standard deviation amounts 1% of the stationary value of the output signal.

For proving the validity of the approximations they are compared with the statistics of 50 test series.

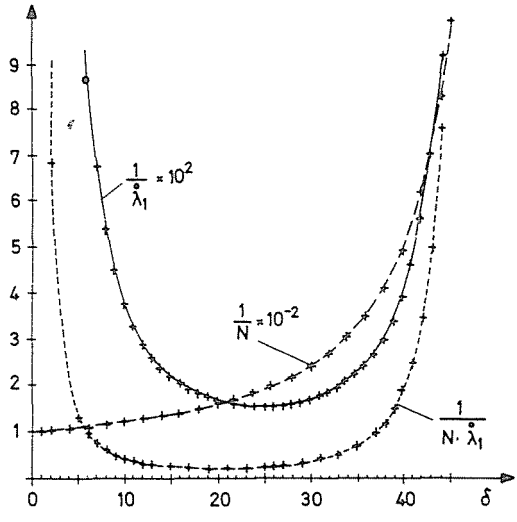

Fig. 6. The degree of linear dependence $\frac{1}{\lambda_1}$ and the inverse of the degree of the overdetermination N

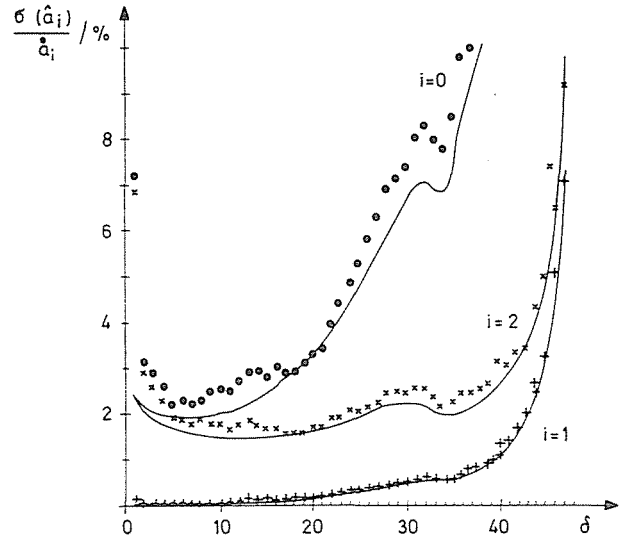

Fig. 7. The normalized standard deviations of the parameters a_i, analytical approximation (—) and statistics of test series (. . ., × × ×, + + +)

According to (4.12) the bias of the estimation vanishes, a similar result shows the computer simulation for all admissible δ values. The variances of the estimated parameters are described by the covariance matrix (4.15). In Fig. 6 the dependence of the important factors $\frac{1}{N}$ and $\frac{1}{\lambda_1}$ from the sampling factor δ is plotted. The number N of equations and therefore, the degree of overdetermination of the equation system decreases with increasing δ. The degree of linear dependence given by $\frac{1}{\lambda_1}$ shows a relatively broad minimum. In Fig. 7 the normalized standard deviations of the parameters $\hat{a}_i$ can be compared. The approximations are in good agreement with the test series. For the uncorrelated domain ($33 \leq \delta \leq 48$) the approximations are computed by Eq. (4.18). We can see that very small and great values for the effective sampling time $\delta \cdot \Delta$ in the autoregressive equation should be avoided. In this case a sampling factor δ around 5. . .15 is preferable.

References

1. GRAUPE, D. *et al.*: A Comparative Analysis of Various Least-Squares Identification Algorithms. Automatica, *16*, No. 6, pp. 663—681. (1980).
2. HAYKIN, S. (Ed.): Nonlinear Methods of Spectral Analysis. Springer-Verlag, Berlin—Heidelberg—New York, 1980.
3. KOOPMANS, T.: Linear Regression Analysis of Economic Time Series. De Erven F. Bohn N. V., Harlem, 1936.
4. KRONMÜLLER, H.: Konsistente Parameterschätzung mit orthogonaler Regression. Bericht vom Institut für Prozeßmeßtechnik und Prozeßleittechnik, Universität Karlsruhe, 1978.
5. KUHNERT, F.: Pseudoinverse Matrizen und die Methode der Regularisierung. Teubner Verlagsgesellschaft, Leipzig, 1976.
6. LEVIN, M. J.: Estimation of a System Pulse Transfer Function in the Presence of Noise. IEEE Trans. on Automatic Control vol. AC—*9*, pp. 229—235, 1964.
7. PAPOULIS, A.: Probability, Random Variables and Stochastic Processes. McGraw-Hill, Kogakusha, 1965.
8. PROSS, D.: Näherungen für die Verzerrung und die Varianz einer Klasse autoregressiver Schätzverfahren. Dissertation, Universität Karlsruhe, 1984.
9. SORENSON, W.: Parameter Estimation. Dekker, New York, 1980.
10. STÖCKLE, J.: Estimation of Signal Parameters in Exogenous Process Models. Periodica Polytechnica, this issue.
11. STRANG, G.: Linear Algebra and its Applications. Academic Press, New York, San Francisco, London, 1980.

Dieter PROSS Institut für Prozeßmeßtechnik und
Prozeßleittechnik,Universität Karlsruhe, Hertzstr. 16.
D-7500 Karlsruhe,
German Federal Republic