On the Euclidean Algorithm Applied to Polynomials of Models Obtained by Least Squares Identification

Söderström, Torsten

1973

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Söderström, T. (1973). *On the Euclidean Algorithm Applied to Polynomials of Models Obtained by Least Squares Identification*. (Research Reports TFRT-3108). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

TFRT-3108

ON THE EUCLIDEAN ALGORITHM APPLIED TO POLYNOMIALS OF MODELS OBTAINED BY LEAST SQUARES IDENTIFICATION

T. SÖDERSTRÖM

Report 7328(C) November 1973
Lund Institute of Technology
Division of Automatic Control

ON THE EUCLIDEAN ALGORITHM APPLIED TO POLYNOMIALS OF MODELS OBTAINED BY LEAST SQUARES IDENTIFICATION

T. Söderström

Abstract

In identification of systems with noise the least squares method is sometimes applied to models of high order. After the identification it can be desired to reduce the obtained model. One way to do this reduction is to use the Euclidean algorithm. This method was examined by Burström, who found in numerical examples that this reduction gives consistent estimates provided that the input signal is white noise and that the model order is chosen properly. The purpose of this report is to prove this fact.

TABLE OF CONTENTS

I. INTRODUCTION

The least squares (LS) method is a well-known method for identification of linear dynamic systems. It is easy to apply, but the estimates are not consistent if the output signal is corrupted with noise of a general covariance function. A way to overcome this difficulty is proposed e.g. in Åström - Eykhoff (1971). They suggest a high model order and call the method "the repeated least squares method."

Consider a system given by

$$A(q)y(t) = B(q)u(t) + v(t) \tag{1.1}$$

where $y(t)$ denotes the output signal, $u(t)$ the input signal, and $v(t)$ is noise. It is assumed that $v(t)$ is a stationary, ergodic process. It may have an arbitrary covariance function. The notation q means the forward shift operator and

$$A(q) = q^n + a_1 q^{n-1} + \ldots + a_n$$

$$B(q) = b_1 q^{n-1} + \ldots + b_n$$

The two polynomials can always be chosen as relatively prime.

It is assumed that the least squares model of the true system is of the form

$$\hat{A}(q)y(t) = \hat{B}(q)u(t) + \varepsilon(t) \tag{1.2}$$

where

$$\hat{A}(q) = q^m + \hat{a}_1 q^{m-1} + \ldots + \hat{a}_m$$

$$\hat{B}(q) = \hat{b}_1 q^{m-1} + \ldots + \hat{b}_m$$

It is well known that the LS method gives consistent estimates only if $v(t)$ is white. The repeated LS method means that m is chosen larger than n (by repeated trials with increase of m) and in succesful cases

$$\hat{A}(q) \approx A(q)L(q) \qquad (1.3)$$

$$\hat{B}(q) \approx B(q)L(q)$$

is fulfilled approximately for some polynomial $L(q)$ of degree $m-n$. In Åström - Eykhoff (1971) an illustrative example is given.

It may often be desirable to re-estimate $A(q)$ and $B(q)$ from $\hat{A}(q)$ and $\hat{B}(q)$. Since (1.3) is not exactly true, the uncertainties of $\hat{a}_1 \ldots \hat{b}_m$ must be taken into account. In Söderström (1973) three different ways to do this re-estimation are discussed.One of these ways is to use the Euclidean algorithm, but to test if the remainder polynomial is zero in a statistical sense in every step. This way was examined in detail in Burström (1973), who has shown that it is hard (or impossible) to give reliable test quantities for determining n in this way. He also found by experience that provided n was chosen correctly in the re-estimation, the inequality $m \geq 3n-1$ held, the input signal was white noise and the asymptotic expressions of $\hat{A}(q)$ and $\hat{B}(q)$ were used, then the correct $A(q)$ and $B(q)$ were obtained.

The purpose of this report is to theoretically prove that this experience of the Euclidean algorithm is generally true. Asymptotic expressions for the parameter estimates will be used throughout the report.

II. NOTATIONS AND PRELIMINARIES

In order to simplify the notations, the degree of a polynomial will sometimes be used as a subscript.
Let

$$A_m(z) = a_o z^m + a_1 z^{m-1} + \ldots + a_m$$

Then $d(A) = d(A_m) = m$ will be used as notations of the degree.

The following well-known lemma will be useful

Lemma 2.1

Assume that two polynomials A_m and B_n are given and that $m \geq n$, $b_o \neq 0$. Then there are two unique polynomials C_{m-n} and D_{n-1} such that

$$A_m = C_{m-n}B_n + D_{n-1} \tag{2.1}$$

(which means $A_m(z) = C_{m-n}(z)B_n(z) + D_{n-1}(z)$ for all z). If $n = 0$, D_{n-1} is to be interpreted as zero.

In the forthcoming analysis the following lemma, proved in Söderström (1973), will also be of vital importance.

Lemma 2.2

Consider the system (1.1) and assume that u(t) is white noise, independent of {v(t)}. Consider the LS model (1.2) and assume $m \geq n$. Using the asymptotic values of $\hat{A}_m(z)$ and $\hat{B}_{m-1}(z)$ the following is true

$$d(\hat{A}_m B_{n-1} - A_n \hat{B}_{m-1}) = n - 1 \tag{2.2}$$

It is shown in Söderström (1973) that Lemma 2.2 can be interpreted in the following way. It means that the descrete time impulse responses of the transfer functions $B(q)/A(q)$ and $\hat{B}(q)/\hat{A}(q)$ coincide in the first m points.

III. MAIN RESULT

The reader is assumed to be familiar with the Euclidean algorithm applied to polynomials. (A short description is given in the next section). In the present situation this algorithm is applied to $\hat{A}_m$ and $\hat{B}_{m-1}$. It is assumed that the procedure is interrupted when a remainder polynomial of degree m-n is obtained. Multiply this polynomial with a constant so that it becomes monic and denote it afterwards by C_{m-n}. Then A_n and B_{n-1} are estimated (re-estimated) by polynomial divisions of $\hat{A}_m$ resp. $\hat{B}_{m-1}$ with C_{m-n}, i.e. given by

$$d(\hat{A}_m - \hat{\hat{A}}_n C_{m-n}) \leq m - n - 1$$

$$d(\hat{B}_{m-1} - \hat{\hat{B}}_{n-1} C_{m-n}) \leq m - n - 1 \qquad (3.1)$$

Theorem 3.1

Consider the system (1.1) and assume that u(t) is white noise. Then the Euclidean algorithm as described above will give consistent estimates, i.e.

$$\hat{\hat{A}}_n = A_n, \; \hat{\hat{B}}_{n-1} = B_n \qquad (3.2)$$

if

$$m \geq 3n - 1 \qquad (3.3)$$

The proof of the theorem is rather long and technical. It is given in the next section.

Remark 1. It is stated that (3.3) is a sufficient condition. Due to the experience by Burström (1973) it seems to be a necessary condition as well.

Remark 2. In Söderström (1973) an alternate method for the re-estimation of A(q) and B(q) is discussed. For this method it has been proved that $m \geq 2n$ replaces (3.3) as a sufficient condition using the same assumptions.

IV. PROOF OF THE MAIN RESULT

The major part of the proof of theorem 3.1 will be given as four lemmas.

Lemma 4.1

With the assumptions of Lemma 2.2 there are unique L_{m-n}, $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$ such that

$$\hat{A}_m = L_{m-n}A_n + \tilde{A}_{n-1}$$

$$\hat{B}_{m-1} = L_{m-n}B_{n-1} + \tilde{B}_{n-2} \tag{4.1}$$

If $n = 1$, $\tilde{B}_{n-2}$ is to interpreted as zero

Proof: From $\hat{A}_m$ and A_n the polynomials L_{m-n} and $\tilde{A}_{n-1}$ are uniquely given (Lemma 2.1). From Lemma 2.2 it follows that

$$\hat{A}_m B_{n-1} - A_n\hat{B}_{m-1} = P_{n-1} \tag{4.2}$$

for some polynomial P_{n-1}.

Then

$$\hat{B}_{m-1} - L_{m-n}B_{n-1} = \frac{\hat{A}_m B_{n-1} - P_{n-1}}{A_n} - L_{m-n}B_{n-1} =$$

$$= \frac{(L_{m-n}A_n + \tilde{A}_{n-1})B_{n-1} - P_{n-1} - L_{m-n}B_{n-1}A_n}{A_n} =$$

$$= \frac{\tilde{A}_{n-1}B_{n-1} - P_{n-1}}{A_n}$$

However, this expression is a polynomial and not a fraction. It turns out that the degree of the polynomial is $d(\tilde{A}_{n-1}) + d(B_{n-1}) - d(A_n) = n - 2$.

Thus $\hat{B}_{m-1} = L_{m-n}B_{n-1} + \tilde{B}_{n-2}$. □

Assume for a moment that $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$ are dropped in (4.1). Then the Euclidean algorithm will give L_{m-n} as a common factor of $\hat{A}_m$ and $\hat{B}_{m-1}$. Then $\hat{\hat{A}}_n$ and $\hat{\hat{B}}_{n-1}$ can be computed from (3.1) with $C_{m-n} = L_{m-n}$. Obviously the result is $\hat{\hat{A}}_n = A_n$, $\hat{\hat{B}}_{n-1} = B_{n-1}$.

The following analysis will show that this result of the re-estimation will not be influenced by $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$ if m is chosen properly.

In the use of the Euclidean algorithm, which can be described as successive applications of Lemma 2.1, it is possible that a new polynomial (corresponding to D_{n-1} is the lemma) has a zero as its first coefficient. In such a case the degree must be considered as one unit smaller. The consideration of this degeneration leads to rather technical calculations.

In order to describe the procedure on $\hat{A}_m$ and $\hat{B}_{m-1}$ it is first considered how the Euclidean algorithm will work on A_n and B_{n-1}. The latter case is described by two sequences of polynomials $\{Q^i\}_o^p$ and $\{K^i\}_{i=1}^{p-1}$ which are defined by

$$Q_n^o = A_n, \qquad Q_{n-1}^1 = B_{n-1}$$

$$Q^i: Q^{i-2} = Q^{i-1} K^{i-1} + Q^i \qquad i = 2 \dots p$$

Put further $q_i = d(Q^i)$ and $k_i = d(K^i)$. If no degeneration occurs, then $k_i = 1$ for all i and $q_i = n - i$. Since A_n and B_{n-1} are relatively prime it is obvious that the iterations will finally give a non-zero constant, which is denoted Q^p. If no degenerations occur, p = n. In the more general case k_i may be larger than 1 and

$$q_i = q_{i-1} - k_i$$

holds, (k_p is formally defined as $q_p - q_{p-1}$). According to the definition of p, $q_p = 0$.

The number of degenerations is $\sum_{i=1}^{p}(k_i-1) = n - p$.

Consider now the Euclidean algorithm applied to $\hat{A}_m$ and $\hat{B}_{m-1}$. It can then be described by the sequences

$(S^i)_{i=0}^{\hat{p}}$, $(\hat{K}^i)_{i=0}^{\hat{p}}$ defined by

$$S^o = \hat{A}_m, \quad S^1 = \hat{B}_{m-1}$$

$$S^i: \quad S^{i-2} = S^{i-1}\,\hat{K}^{i-1} + S^i \qquad i = 2 \ldots \hat{p}$$

As described earlier, it is desirable that $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$ have "small" influence. For this reason it is relevant to ask when $\hat{K}^i = K^i$. Consider especially the case i = 1 and utilize (A.1). Then

$$(L_{m-n}A_n + \tilde{A}_{n-2}) = (L_{m-n}B_{n-1} + \tilde{B}_{n-2})\,\hat{K}^1 + S^2$$

which must be compared with

$$A_n = B_{n-1}K^1 + Q^2$$

The polynomial K_1, which is of degree 1, is uniquely determined by a_1, b_1, and b_2. In the same way K^i is uniquely given by $\hat{a}_1$, $\hat{b}_1$, and $\hat{b}_2$. Clearly if $\tilde{A}_{n-2}$ and $\tilde{B}_{n-2}$ do not affect $\hat{K}^1$ then $\hat{K}^1 = K^1$ must hold. This is the case if $\deg(\tilde{A}_{n-1}) \leq \deg(\hat{A}_m) - 2$ and, $\deg(\tilde{B}_{n-2}) \leq \deg(\hat{B}_{m-1}) - 2$, i.e. $m \geq n + 1$.

This idea will now be generalized. Assume that $\hat{K}^i = K^i$ for all i. Consider first the problem of solving (3.1) for $\hat{\hat{A}}_n$ and $\hat{B}_{n-1}$. The polynomial C_{m-n} can be written as $L_{m-n} + \tilde{C}$ where $\tilde{C}$ includes the effects due to $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$. With use of (4.1) the equations (3.1) are rewritten as

$$d(A_n L_{m-n} + \tilde{A}_{n-1} - \hat{\hat{A}}_n(L_{m-n} + \tilde{C})) \leqslant m - n - 1$$

$$d(B_{n-1} L_{m-n} + \tilde{B}_{n-2} - \hat{B}_{n-1}(L_{m-n} + \tilde{C})) \leqslant m - n - 1$$

These equations are assumed to have the unique solution (3.2) which gives the following necessary and sufficient condition.

$$d(\tilde{A}_{n-1} - A_n \tilde{C}) \leqslant m - n - 1$$

$$d(\tilde{B}_{n-2} - B_{n-1} \tilde{C}) \leqslant m - n - 1$$

which is equivalent to

$$n + \deg(\tilde{C}) \leqslant m - n - 1$$

or

$$\deg(\tilde{C}) \leqslant m - 2n - 1 \qquad (4.3)$$

The condition (4.3) means that the first 2n coefficients of C_{m-n} defined in section 3 coincide with corresponding coefficients of L_{m-n}. For this reason it must be assumed that S^i and $Q^i L_{m-n}$ coincide in the beginning. In order to carry out these conditions precisely introduce the sequence $\{R^i\}_{i=0}^{p}$, which will describe the effect of $\tilde{A}_{n-1}$ and $\tilde{B}_{n-2}$ on S^i, by

$$R^o = \tilde{A}_{n-1}, \; R^1 = \tilde{B}_{n-2}$$

$$R^i = R^{i-2} - R^{i-1} K^{i-1}$$

and put $r_i = d(R^i)$

<u>Lemma 4.2</u>

The sequence $\{S_i\}_{i=0}^{p}$ satisfies

$$S^i = Q^i L + R^i \qquad 0 \leqslant i \leqslant p \qquad (4.4)$$

if

$$k_i < \min(q_i+m-n-r_i,\ q_{i-1}+m-n-r_{i-1}) \qquad 1 \leq i \leq (p-1) \qquad (4.5)$$

Proof: By definitions (4.4) holds for $i = 0$ and 1. Assume that (4.4) is satisfied for $i = 1, \ldots (k-1)$. Consider the situation for $i = k$. The definition of S^k and (4.4) imply

$$(Q^{k-2}_{q_{k-2}} L_{m-n} + R^{k-2}_{r_{k-2}}) = (Q^{k-1}_{q_{k-1}} L_{m-n} + R^{k-1}_{r_{k-1}}) \hat{K}^{k-1} + S^k$$

According to the condition (4.5) with $i = k-1$ it is concluded that the polynomial $\hat{K}$ is uniquely given by $Q^{k-2}L$ and $Q^{k-1}L$, which means that $\hat{K}^{k-1} = K^{k-1}$. Thus

$$S^k = (Q^{k-2}L+R^{k-2}) - (Q^{k-1}L+R^{k-1})\ K^{k-1}$$

$$= (Q^{k-2}-Q^{k-1}K^{k-1})L + (R^{k-2}-R^{k-1}K^{k-1})$$

$$= Q^kL + R^k$$

which finishes the proof. □

When (4.4) holds it means that $S^p = Q^p_o L_{m-n} + R^p$ and that $C_{m-n} = S^p/Q^p_o$. Further, the polynomial $\tilde{C}$ becomes $\tilde{C} = R^p/Q^p_o$ and thus $\deg(\tilde{C}) = r_p$. This means that (4.3) and (4.5) can be rewritten as

$$r_i < q_i + m - n - k_{i+1} \qquad 0 \leq i \leq p-2$$

$$r_i < q_i + m - n - k_i \qquad 1 \leq i \leq p-1$$

$$r_p < m - 2n$$

Summarizing the discussion so far, it has been proved that (4.6) implies that theorem 3.1 is true. It remains to be proved that (3.3) implies (4.6).

Lemma 4.3

Consider the sequences $\{R^i\}$and $\{Q^i\}$as defined above. Then

$$R^iQ^{i+1} - R^{i+1}Q^i = (-1)^i\ P_{n-1} \qquad 0 \leq i \leq p-1 \qquad (4.7)$$

where P_{n-1} is given by (4.2)

Proof: With use of (4.1) and (4.2)

$$P_{n-1} = \hat{A}B - A\hat{B} = (LA+\tilde{A})B - A(LB+\tilde{B}) =$$

$$= \tilde{A}B - A\tilde{B} = R^oQ^1 - Q^oR^1$$

which proves (4.7) for $i = 0$.

Assume now that (A.6) is true for $i = k - 1$, i.e.

$$R^{k-1}Q^k - Q^{k-1}R^k = (-1)^{k-1}P$$

Then using the definitions of R^{k+1} and Q^{k+1} and assuming $k \leq p-1$

$$R^kQ^{k+1} - Q^kR^{k+1} = R^k(Q^{k-1}-Q^kK^k) -$$

$$- Q^k(R^{k-1}-R^kK^k) = R^kQ^{k-1} - Q^kR^{k-1} = (-1)^kP$$

which proves the lemma □

Lemma 4.4

The numbers r_i satisfy

$$r_i \leq n - 2 \qquad 1 \leq i \leq p \qquad (4.8)$$

Proof: First (4.7) is rewritten as $R^{i+1}Q^i = R^iQ^{i+1} - (-1)^iP_{n-1}$ which implies

$$r_{i+1} \leq \max(r_i+q_{i+1},\ n-1) - q_i \qquad (4.9)$$

By the definition of R^1, (4.8) is true for $i = 1$. Assume that (4.8) is true for $1 \leq i \leq k < p-1$. Since $q_i = n - \sum_1^i k_j$ (4.9) implies

$$r_{k+1} \leq \max(r_k+q_{k+1}, n-1) - q_k$$

$$\leq \max(n-2+q_{k+1}, n-1) - q_k = n - 2 + q_{k+1} - q_k < n - 2$$

Thus (4.8) holds for $1 \leq i \leq (p-1)$. Finally

$$r_p = \max(r_{p-1}+q_p, n-1) - q_{p-1}$$

$$\leq \max(n-2+0, n-1) - q_{p-1} = n - 1 - k_p \leq n - 2$$

□

With use of (4.8) it is possible to rewrite (4.6) into a new condition involving only m and n, namely (3.3).

The sufficient conditions (4.6) are now rewritten as follows

$$r_o < q_o + m - n - k_1$$

$$r_i < q_i + m - n - \max(k_i, k_{i+1}) \qquad 1 \leq i \leq p - 2 \qquad (4.10)$$

$$r_{p-1} < q_{p-1} + m - n - k_{p-1}$$

$$r_p < m - 2n$$

With use of the definitions of R^o, Q^o, and K^1 the first inequality turns out to be

$$n - 1 < n + m - n - 1$$

or $m > n$. The remaining inequalities of (4.10) are satisfied if (use lemma 4.4)

$$n - 2 < q_i + m - n - \max(k_i, k_{i+1}) \qquad 1 \leq i \leq p - 1 \qquad (4.11)$$

$$n - 2 < m - 2n$$

The last inequality of (4.11) can be written as

$$m \geq 3n - 1$$

which is precisely (3.3). The first sequence of inequalities of (4.11) is rewritten as

$$m > 2n - 2 - (n - \sum_{1}^{i} k_j) + \max(k_i, k_{i+1}) \qquad 1 \leq i \leq p - 1 \qquad (4.12)$$

The right hand side of (4.12) satisfies

$$2n - 2 - (n - \sum_{1}^{i} k_j) + \max(k_i, k_{i+1})$$

$$= n - 2 + \max(\sum_{1}^{i} k_j + k_i, \sum_{1}^{i+1} k_j)$$

$$\leq n - 2 + \max(\sum_{1}^{p} k_j + k_i, \sum_{1}^{p} k_j)$$

$$= 2n - 2 + k_i \leq 3n - 2$$

This means that (4.12) is implied by the condition (3.3). With this observation the proof of theorem 3.1 is finished. □

REFERENCES

Burström, E. (1973)
Förkortning av Gemensamma Faktorer i Skattade Överföringsfunktioner. (In Swedish). Master thesis RE 124, Division of Automatic Control, Lund Institute of Technology.

Söderström, T. (1973)
Test of Common Factors in Identified Models. Application to the Generalized Least Squares Method. Report 7328, Division of Automatic Control, Lund Institute of Technology.