

Identification Aspects of Ritt's Algorithm for Discrete-Time Systems

Christian Lyzell, Torkel Glad, Martin Enqvist, Lennart Ljung

Division of Automatic Control

E-mail: lyzell@isy.liu.se, torkel@isy.liu.se,
maren@isy.liu.se, ljung@isy.liu.se

3rd April 2009

Report no.: LiTH-ISY-R-2891

Accepted for publication in 15th IFAC Symposium on System Identification, Saint-Malo, France

Address:

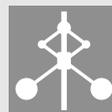
Department of Electrical Engineering

Linköpings universitet

SE-581 83 Linköping, Sweden

WWW: <http://www.control.isy.liu.se>

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET



Abstract

In system identification, the challenge of parameter estimation often lies in solving a non-convex optimization problem. In many cases, this implies that it is difficult to guarantee that the global optimum will be found. The tools of differential algebra, for example, Gröbner bases and Ritt's algorithm, have turned out to be quite useful when dealing with certain nonlinear model structures. Some examples of successful applications are the determination of controllability, observability and global identifiability of these model structures. In this paper, difference algebraic techniques, which mimics the differential algebraic methods, will be presented. Besides making it possible to handle discrete-time systems, this opens up the possibility of dealing with noise. It turns out that the classical instrumental variables method plays a role.

Keywords: System identification, Identifiability, Ritt's algorithm.

Identification Aspects of Ritt's Algorithm for Discrete-Time Systems

C. Lyzell* T. Glad* M. Enqvist* L. Ljung*

* *Division of Automatic Control, Department of Electrical Engineering,
Linköpings Universitet, SE-581 83, Linköping, Sweden.
(E-mail: {lyzell,torkel,maren,ljung}@isy.liu.se)*

Abstract: In system identification, the challenge of parameter estimation often lies in solving a non-convex optimization problem. In many cases, this implies that it is difficult to guarantee that the global optimum will be found. The tools of differential algebra, for example, Gröbner bases and Ritt's algorithm, have turned out to be quite useful when dealing with certain nonlinear model structures. Some examples of successful applications are the determination of controllability, observability and global identifiability of these model structures. In this paper, difference algebraic techniques, which mimics the differential algebraic methods, will be presented. Besides making it possible to handle discrete-time systems, this opens up the possibility of dealing with noise. It turns out that the classical instrumental variables method plays a role.

Keywords: System identification, Identifiability, Ritt's algorithm.

1. INTRODUCTION

The identification of linear and nonlinear parametric models is an important part of system identification. Unfortunately, even seemingly simple model structures can yield difficult challenges, in terms of non-convex optimization problems, when estimating the parameters. In some cases, it is therefore difficult to guarantee that the global optimum will be found. For methods like the *Prediction-Error Method* (PEM), an essential preliminary step is to find an initial parameter estimate lying in the region of attraction of the global optimum.

For example, consider the problem of estimating the parameter θ in

$$y(t) = \theta u(t) + \theta^2 u(t-1) + e(t), \quad (1)$$

where y is the output, u is the input, and e is the noise. If one formulates the estimation problem as minimizing the squared prediction error with respect to the parameter θ , one notices that the optimization problem becomes non-convex, even for such a simple model structure.

Let us, for example, time shift (1):

$$y(t-1) = \theta u(t-1) + \theta^2 u(t-2) + e(t-1). \quad (2)$$

By examining the equations, we see that by multiplying (1) with $u(t-2)$ and (2) with $u(t-1)$, and then subtracting the results, it holds that

$$\begin{aligned} u(t-2)y(t) - u(t-1)y(t-1) \\ = (u(t-2)u(t) - u^2(t-1))\theta \\ + u(t-2)e(t) - u(t-1)e(t-1). \end{aligned} \quad (3)$$

This equation is linear in θ and thus, the original non-convex problem can be written as a linear regression, i.e. the estimation problem becomes convex.

The calculations above are inspired by the methods of differential algebra, which have been successfully applied to a variety of nonlinear continuous-time problems. For example, in Ljung and Glad (1994), these methods were used to prove quite general identifiability results. The main result was that a nonlinear model structure is globally identifiable if and only if it can be written as a linear regression. In this paper, we will discuss similar results in discrete-time.

Also, if one can generalize the methods of differential algebra to work for discrete-time systems, differentiation of noise signals will no longer be an issue and the noise can be manipulated as any other signal. This will lead to some interesting system identification problems since the noise signal will be transformed in a quite complex manner.

Attempts to mimic Ritt's algorithm in discrete-time have been made by Kotsios (2001), where the δ -operators are introduced and different products of these are discussed. The aim in this paper is to generalize Ritt's algorithm with a minimum amount of changes compared to the continuous-time case.

The basic methods of differential algebra are given by Ritt (1950), see also Kolchin (1973) and Seidenberg (1956). An introduction to differential algebra is given in Fliess and Glad (1993). Algorithmic aspects are discussed in Glad (1992) and in Diop (1989, 1991).

In Section 2, the basic algebraic concepts used in this paper are presented and in Section 3 the discrete-time Ritt's algorithm is given. In Section 4 some results on identifiability are given, while Section 5 discusses some aspect of system identification when using Ritt's algorithm. Finally, conclusions and future work are given in Section 6.

2. BASIC CONCEPTS

As discussed above, we are interested in systems described by polynomials in time shifted variables. The time displacement will be denoted by dots or the order of the displacement in parenthesis:

$$u = u(t), \dot{u} = u(t+1), \ddot{u} = u(t+2), u^{(3)} = u(t+3), \dots$$

Alternatively we might use displacements backwards in time

$$u = u(t), \dot{u} = u(t-1), \ddot{u} = u(t-2), u^{(3)} = u(t-3), \dots$$

Note that we will always denote $u^{(\sigma)}$ to be a positive time displacement with respect to $u^{(\nu)}$ if $\nu > \sigma$ even if the underlying displacement is backwards in time as in (2). A fundamental concept for the algorithmic aspects of differential algebra is *ranking*. This is a total ordering of all variables and their derivatives. In the discrete-time case it is a total ordering of all time shifted variables. Examples involving two variables are

$$u < y < \dot{u} < \dot{y} < \ddot{u} < \ddot{y} < \dots \quad (4)$$

and

$$u < \dot{u} < \ddot{u} < \dots < y < \dot{y} < \ddot{y} < \dots \quad (5)$$

where $<$ denotes "is ranked lower than". Any ranking is possible provided that it satisfies the conditions:

$$u^{(\mu)} < u^{(\mu+\sigma)} \quad (6)$$

$$u^{(\mu)} < y^{(\nu)} \Rightarrow u^{(\mu+\sigma)} < y^{(\nu+\sigma)} \quad (7)$$

for all variables u and y , all nonnegative integers μ and ν , and all positive integers σ . The highest ranked time shifted variable in a (possibly time shifted) polynomial is called the *leader*.

The ranking of time shifted variables gives a ranking of time shifted difference polynomials. They are simply ranked as their leaders. If they have the same leader, they are considered as polynomials in their leader and the one of lowest degree is ranked lower.

The degree of a variable v in a polynomial A is denoted by $\deg_v A$.

Let A, B be two polynomials and let A have the leader v . Then B is said to be *reduced* with respect to A if there is no positive time displacement of v in B and if $\deg_v B < \deg_v A$.

A set $\{A_1, \dots, A_p\}$ of polynomials is called *auto-reduced* if all the A_i are pairwise reduced with respect to each other. Normally, auto-reduced sets are ordered so that A_1, \dots, A_p are in increasing rank.

Auto-reduced sets are ranked as follows. Let $\mathbf{A} = \{A_1, \dots, A_r\}$ and $\mathbf{B} = \{B_1, \dots, B_s\}$ be two ordered auto-reduced sets. \mathbf{A} is ranked lower if either there is an integer k , $1 \leq k \leq \min(s, r)$ such that

$$\text{rank } A_j = \text{rank } B_j, \quad j = 1, \dots, k-1 \quad (8)$$

$$\text{rank } A_k < \text{rank } B_k \quad (9)$$

or else if $r > s$ and

$$\text{rank } A_j = \text{rank } B_j, \quad j = 1, \dots, s \quad (10)$$

A *characteristic set* for a given set of time shifted polynomials is an auto-reduced subset such that no other auto-reduced subset is ranked lower.

The *separant* S_A of a polynomial A is the partial derivative of A with respect to the leader, while the *initial* I_A is the coefficient of the highest power of the leader in A .

If a time shifted polynomial f is not reduced with respect to another time shifted polynomial g , then a reduced polynomial can be constructed using the following series of lemmas.

Lemma 1. Let g be a polynomial with leader u and let f be a polynomial containing $u^{(\sigma)}$. Then there exist polynomials R and Q such that

- $\deg_{u^{(\sigma)}} R < \deg_{u^{(\sigma)}} f$.
- Every solution of $f = 0, g = 0$ is also a solution of $R = 0, g = 0$.
- Every solution of $R = 0, g = 0$ with $Q \neq 0$ is also a solution of $f = 0, g = 0$.

Proof. Let μ, ν be the degrees of f and $g^{(\sigma)}$ respectively as polynomials in $u^{(\sigma)}$. Then, if $\mu \geq \nu$, polynomial division gives

$$Qf = \bar{Q}g^{(\sigma)} + R$$

with $\deg_{u^{(\sigma)}} R < \nu \leq \mu$. If, on the other hand $\mu < \nu$ then

$$\bar{Q}g^{(\sigma)} = Qf + R$$

with $\deg_{u^{(\sigma)}} R < \mu$ □

Lemma 2. Let g be a polynomial with leader u and let f be a polynomial containing $u^{(\sigma)}$. Then there exist polynomials R and Q such that

- R does not contain $u^{(\sigma)}$.
- Every solution of $f = 0, g = 0$ is also a solution of $R = 0, g = 0$.
- Every solution of $R = 0, g = 0$ with $Q \neq 0$ is also a solution of $f = 0, g = 0$.

Proof. This follows from repeated application of Lemma 1. □

Lemma 3. Let f be a polynomial that is not reduced with respect to the polynomial g . Then there exist polynomials R and Q such that

- R is reduced with respect to f .
- Every solution of $f = 0, g = 0$ is also a solution of $R = 0, g = 0$ for all t .
- Every solution of $R = 0, g = 0$ with $Q \neq 0$ is also a solution of $f = 0, g = 0$.

Proof. If f is not reduced with respect to g , then either f contains some positive time displacement of the leader u of g or else f contains u to a higher power.

In the latter case one could perform a pseudo-division of f by g getting

$$Qf = \bar{Q}g + R \quad (11)$$

for some polynomials Q and \bar{Q} , where R is reduced with respect to g .

In the former case f contains $u^{(\sigma)}$ for some positive integer σ and Lemma 2 can be used (possibly several times), followed by (11). The final Q will be the product of the Q 's generated in the individual steps. □

The following property is important for the finiteness of the algorithms that we are going to present.

Proposition 4. A sequence of time displacements of a finite number of variables, each one ranked lower than the preceding one, can only have finite length.

Proof. Let y_1, \dots, y_p denote all the variables whose time displacement appear anywhere in the sequence. For each y_j let σ_j denote the order of the first appearing time shift. There can then be only σ_j lower time displacements of y_j in the sequence. The total number of elements is thus bounded by $\sigma_1 + \dots + \sigma_p + p$. \square

A direct consequence is

Proposition 5. A sequence of characteristic sets containing a finite number of variables, each one ranked lower than the preceding one, can only have finite length.

3. THE ALGORITHM

Consider to begin with a system of the form

$$f_1 = 0, \dots, f_n = 0 \quad (12)$$

where the f_i are polynomials in the variables

$$y, u, x, \theta,$$

and their time-shifts. The basic elimination algorithm is now presented in Algorithm 1.

Algorithm 1 Ritt-Seidenberg

Input: Two sets

$$F = \{f_1, \dots, f_n\}, \quad G = \emptyset$$

where the f_i are polynomials.

- (1) Compute a characteristic set $\mathbf{A} = \{A_1, \dots, A_p\}$ of F .
 - (2) If $F \setminus \mathbf{A}$ is non-empty, then go to Step 4.
 - (3) Add S_A, I_A for all $A \in \mathbf{A}$ to G .
Finished. Output: (F, G) .
 - (4) Let f_k be the highest unreduced (with respect to \mathbf{A}) element of F . Then use Lemma 3 to get an R and a Q . Add Q to G .
 - (5) If $R = 0$, then $F := F \setminus \{f_k\}$ and go to Step 1.
 - (6) Set $F := (F \setminus \{f_k\}) \cup \{R\}$ and go to Step 1.
-

From here on, the Ritt-Seidenberg algorithm will be referred to as Ritt's algorithm.

Proposition 6. Ritt's algorithm will reach the point marked "Finished" after a finite number of steps.

Proof. The only possible loop is via Step 5 or Step 6 to Step 1. This involves either the removal of a polynomial or its replacement with one that is reduced with respect to \mathbf{A} or has its highest unreduced time displacement removed. If R is reduced, then it is possible to construct a lower auto-reduced set. An infinite loop would thus contradict either Proposition 4 or Proposition 5. \square

Proposition 7. Every solution to the initial set F of the algorithm is also a solution to the final set. Every solution to the final set for which the polynomials of G are nonzero is also a solution of the initial set.

Proof. Follows from a repeated application of Lemma 3. \square

Remark 8. The inclusion of separants and initials in Step 3 might be unnecessary in some applications. It is useful in order to guarantee existence of solutions to the resulting system.

We conclude this section by repeating the example in Section 1, but here the parameter nonlinearity is eliminated using the Ritt's algorithm. The ordering used is

$$u < \dot{u} < \dots < e < \dot{e} < \dots < y < \dot{y} < \dots \quad (13a)$$

$$< \theta_1 < \theta_2 < \dots < \dot{\theta}_1 < \dot{\theta}_2 < \dots \quad (13b)$$

Example 9. Consider the problem of estimating the parameter θ in the model

$$y(t) = \theta u(t) + \theta^2 u(t-1) + e(t), \quad (14)$$

given input and output data. This problem could be solved by defining $\tilde{\theta} = \theta^2$ and solving a least-squares problem, but let us consider using Ritt's algorithm. Define

$$f_1 \triangleq y - u\theta - \dot{u}\theta^2 - e, \quad (15)$$

$$f_2 \triangleq \theta - \dot{\theta}, \quad (16)$$

where the dot operator denotes the backward shift. Let $F \triangleq \{f_1, f_2\}$ and $G = \emptyset$ be the sets of equations and inequations, respectively. The algorithm yields:

Iteration 1. The leaders v_{f_1} and v_{f_2} in F are θ and $\dot{\theta}$, respectively. Since $v_{f_1} < v_{f_2}$, f_1 is reduced with respect to f_2 . The other way around, since $\dot{v}_{f_1} = \dot{\theta}$, f_2 is not reduced with respect to f_1 , i.e. the largest auto-reduced set \mathbf{A} is given by $\{f_1\}$. The division algorithm yields $\bar{Q}_1 f_2 = Q_1 \dot{f}_1 + R_1$, with

$$R_1 = -\dot{y} + \dot{e} + \dot{u}\theta + \ddot{u}\theta\theta,$$

$\bar{Q}_1 = \dot{u} + \ddot{u}\dot{\theta}$, and $Q_1 = 1$. Since R_1 still depends on $\dot{\theta}$, yet another division is needed, i.e. $\bar{Q}_2 f_2 = Q_2 R_1 + R_2$, with

$$R_2 = \dot{y} - \dot{e} - \dot{u}\theta - \ddot{u}\theta^2,$$

$\bar{Q}_2 = -\ddot{u}\theta$, and $Q_2 = 1$. Putting it all together yields

$$\bar{Q} f_2 = Q \dot{f}_1 + R,$$

where $\bar{Q} = \bar{Q}_2 - Q_2 \bar{Q}_1$, $Q = -Q_1 Q_2$, and $R = R_2$. Finally, update G and set $F = \{f_1, f_3\}$ where $f_3 \triangleq R$.

Iteration 2. The leaders in F are now θ and θ , respectively. Thus, $\mathbf{A} = \{f_1\}$ (or $\{f_3\}$) is the characteristic set of F . Division yields $\bar{Q} f_3 = Q f_1 + R$, where

$$R = -\ddot{u}y + \ddot{u}e + \dot{u}\dot{y} - \dot{u}\dot{e} + (\ddot{u}u - \dot{u}^2)\theta,$$

$\bar{Q} = \dot{u}$, and $Q = \ddot{u}$. Update G and set $F := \{f_1, f_4\}$ with $f_4 \triangleq R$.

Thus, one of the equations that Ritt's algorithm find is

$$\ddot{u}y - \dot{u}\dot{y} = (\ddot{u}u - \dot{u}^2)\theta + \ddot{u}e - \dot{u}\dot{e}. \quad (17)$$

The algorithm will continue for another two iterations, such that only one polynomial in F depends on θ , but these steps are not presented here. In this case, the results (3) and (17) coincide. \square

Remark 10. It is worth noting that in the first iteration, two divisions had to be made. This is one of the differences between the discrete-time and the continuous-time Ritt's algorithm. In continuous-time, the derivative always leaves a polynomial which is linear in the leader and only one division is needed in each iteration.

4. IDENTIFIABILITY

In Ljung and Glad (1994), necessary and sufficient conditions for global identifiability of continuous-time model structures were given. Here, the result is generalized for discrete-time systems. First, a definition of identifiability is given.

Definition 11. A model structure \mathcal{M} is said to be *globally identifiable* at θ^* with respect to $\mathcal{D}_{\mathcal{M}}$ if

$$\mathcal{M}(\theta) = \mathcal{M}(\theta^*), \theta \in \mathcal{D}_{\mathcal{M}} \Rightarrow \theta = \theta^*. \quad (18)$$

First, we prove that if Ritt's algorithm finishes with a linear regression, then the original system is globally identifiable.

Proposition 12. If the output of the algorithm contains a polynomial of the form $Q\theta - P$ where Q and P only depend on measured quantities, then θ is globally identifiable, provided that Q has full rank for the measured data.

Proof. Since every solution of the original equations is also a solution of the output equations, it follows that there can only be one value of θ that is consistent with the measured values, provided that Q has full rank. \square

Remark 13. The full rank requirement of Q in Proposition 12 can be interpreted as a condition of *persistence of excitation*, see e.g. Ljung (1999).

Now, if Ritt's algorithm does not yield a linear regression.

Proposition 14. Let the ranking be

$$u() < y() < \theta_1^() < \dots < \theta_m^() < x()$$

where x contains any unmeasured variables. Assume that the output of the algorithm is an autoreduced set of the following form

$$\begin{aligned} p_0(u, \dot{u}, \dots, y, \dot{y}, \dots), \dot{\theta}_1 - \theta_1, \\ p_2(u, \dot{u}, \dots, y, \dot{y}, \dots, \theta_1, \theta_2), \dots, \\ p_m(u, \dot{u}, \dots, y, \dot{y}, \dots, \theta_1, \dots, \theta_m), \dots \end{aligned}$$

and that the separants are included in the G -polynomials. Also assume that a solution exists to this set, such that all inequations are satisfied for all t . Then there are infinitely many θ -values compatible with the u and y of this solution, i.e. the system is not identifiable.

Proof. Fixing the values of u and y in the p_0 polynomials to those of the given solution means that the corresponding equation is always satisfied. θ_1 can now be changed to a new arbitrary constant in the second polynomial. If this change is small enough the remaining equations can now be solved successively for the leader due to the nonvanishing of the separant (the implicit function theorem). \square

5. IDENTIFICATION ASPECTS

In this section we will discuss the use of Ritt's algorithm from a system identification perspective. First we discuss an alternative to elimination theory, i.e. the concept of *over-parametrization*.

5.1 Over-parametrization

The concept of *over-parametrization* can be described as follows: in a regression problem, define each group of nonlinear transformations of the original parameters as a new parameter.

Example 15. Consider the problem of estimating the parameter θ in

$$y(t) = \theta u(t) + \theta^2 u(t-1) + e(t), \quad (19)$$

given input and output data. This problem can be solved by defining $\theta_1 = \theta$ and $\theta_2 = \theta^2$ and solve the linear regression by least-squares, i.e. by using over-parametrization.

Since one is only interested in the estimate $\hat{\theta}_1$, one can just ignore the resulting estimate $\hat{\theta}_2$ or reconcile it with the estimate $\hat{\theta}_1$. \square

Example 16. Once again, consider the problem of estimating the parameter θ in

$$y = \theta u + \theta^2 \dot{u} + e, \quad (20)$$

given input and output data. Time shift (20) to get

$$\dot{y} = \theta \dot{u} + \theta^2 \ddot{u} + \dot{e}, \quad (21)$$

keeping in mind that θ is time independent. Batching the two equations yields the over-parameterized linear regression

$$\begin{pmatrix} y \\ \dot{y} \end{pmatrix} = \underbrace{\begin{pmatrix} u & \dot{u} \\ \dot{u} & \ddot{u} \end{pmatrix}}_{\triangleq \Phi} \begin{pmatrix} \theta \\ \theta^2 \end{pmatrix} + \begin{pmatrix} e \\ \dot{e} \end{pmatrix}. \quad (22)$$

Assuming that the regression matrix Φ is invertible, one can multiply (22) with $(\det \Phi)\Phi^{-1}$ to get

$$\begin{pmatrix} \ddot{u}y - \dot{u}\dot{y} \\ \dot{u}\dot{y} - \dot{u}y \end{pmatrix} = (\ddot{u}u - \dot{u}^2) \begin{pmatrix} \theta \\ \theta^2 \end{pmatrix} + \begin{pmatrix} \ddot{u}e - \dot{u}\dot{e} \\ \dot{u}\dot{e} - \dot{u}e \end{pmatrix}. \quad (23)$$

The first equation is exactly the result of Ritt's algorithm. Thus, there seems to be a connection between over-parametrization and Ritt's algorithm.

5.2 Modifications

It turns out that the Ritt's algorithm goes too far when eliminating the parameter nonlinearities:

Example 17. Consider the problem of estimating the parameters $\theta = (\theta_1, \theta_2)^T$ in

$$y + \theta_1 \dot{y} + \theta_1 \theta_2 \ddot{y} = \theta_2 u + e, \quad (24)$$

where the dot denotes backward shift, given input and output data. Running Ritt's algorithm on (24) yields the linear regression $\eta = \varphi^T \theta + \nu$ where

$$\eta = \begin{pmatrix} uy^{(3)} - \dot{u}\ddot{y} + \ddot{y}\dot{y} - y^{(3)}y \\ \dot{u}\dot{y}\dot{y} - u\dot{y}^2 - \ddot{y}\dot{y}^2 + \dot{y}^2 y \end{pmatrix}, \quad (25a)$$

$$\varphi = \begin{pmatrix} y^{(3)}\dot{y} - \dot{y}^2 & 0 \\ 0 & uy^{(3)}\ddot{y} - \dot{u}\dot{y}^2 + \dot{y}^2\dot{y} - y^{(3)}\ddot{y}y \end{pmatrix}, \quad (25b)$$

$$\nu = \begin{pmatrix} \dot{e}\dot{y} - ey^{(3)} \\ ey^{(3)}\ddot{y} - \dot{e}\dot{y}^2\theta_2 \end{pmatrix}. \quad (25c)$$

Notice that in the expression for θ_1 , there are second order products between the different signals, while in the expression for θ_2 there are exclusively third order powers. This implies that it might be more difficult, in some sense, to estimate the second parameter than the first parameter. Of course, one could change the ordering of the parameters when using the algorithm and get the reverse result.

Now, let us try to do the elimination differently. As before, time shift (24)

$$\dot{y} + \theta_1 \ddot{y} + \theta_1 \theta_2 y^{(3)} = \theta_2 \dot{u} + \dot{e}, \quad (26)$$

keeping in mind that the parameters are constant in time. By multiplying (24) with $y^{(3)}$, (26) with \ddot{y} and subtracting the results, the following holds

$$yy^{(3)} - \dot{y}\ddot{y} = \begin{pmatrix} \dot{y}^2 - \dot{y}y^{(3)} \\ uy^{(3)} - \dot{u}\ddot{y} \end{pmatrix}^T \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + ey^{(3)} - \dot{e}\ddot{y}. \quad (27)$$

This linear regression has similar regressors compared with the ones in (25), where now only second order products

exists. This is due to the fact that Ritt's algorithm is not satisfied until it has one equation for each parameter. So the question that arises is, can Ritt's algorithm be modified to take these things to account. There are some possibilities:

- (1) What enables the calculations above is that all the parameters already occur linearly in the original equation (24). Thus, the only thing left to eliminate is the single parameter nonlinearity. Therefore, if the parameter nonlinearities could be ordered higher than the linear occurrences, then one could stop the algorithm prematurely to receive the linear regression given in (27).
- (2) Another alternative is to stop the algorithm when the first linear regression is decided, i.e. the equation that determines θ_1 . Then one could estimate this parameter and feed it back into the original equation (24) and from there estimate θ_2

Further analysis of possible modifications of Ritt's algorithm, to tailor it for system identification, needs to be made. \square

5.3 Instrumental variables

Using the elimination schemes presented above, Ritt's algorithm and similar methods, yield linear regression forms where the noise is deformed. So when using these equations to estimate the unknown parameters, a tool to deal with these noise transformations is needed. One such tool is the *instrumental variables* (IV) method, see e.g. Söderström and Stoica (1980).

Example 18. Once again, consider the estimation problem presented in Example 17. Depending on the input excitation, this problem could be solved using over-parametrization, see Section 5.1. Despite this, let us try using the result (27).

Trying to solve (27) directly via the least-squares method will yield a biased estimate of the θ_1 parameter, since the first element of the regression vector is correlated with the noise. Thus, instruments, uncorrelated with the noise, need to be chosen for the first parameter.

Now, let us perform a Monte Carlo simulation study of the estimation of the parameters θ via (27), with the least-squares (LS) method and the IV method, respectively. The true parameters are given by $\theta = (-0.1 \ 0.2)^T$. Let the noise and the input be independent white Gaussian processes with zero mean and unit variance. With $M = 100$ Monte Carlo runs and data lengths of 10,000 yields the mean estimates

$$\bar{\theta}_{\text{LS}} = \begin{pmatrix} -0.0750 \\ 0.1999 \end{pmatrix}, \quad \bar{\theta}_{\text{IV}} = \begin{pmatrix} -0.0977 \\ 0.2002 \end{pmatrix}, \quad (28)$$

where \dot{u}^2 has been used as instruments for the first parameter, while the instruments for the second parameter have been chosen as the regressors for that parameter. As predicted, using the LS estimator, a biased estimate is given while the IV estimate appears to be unbiased. \square

Finally, let us consider parameter estimation in state-space equations.

5.4 State-Space Models

So far, only measured signals, except the additive noise, have appeared. Below, the parameter estimation of a nonlinear state-space model is considered. Here, the unknown states need to be eliminated completely, and therefore they are ordered higher than any time shift of the parameters, i.e. the ordering (13) is used.

Example 19. Consider the problem of estimating the parameters $\theta = (\theta_1 \ \theta_2)^T$ in the nonlinear state-space model

$$x_1(t+1) = \theta_1 x_1(t) + \theta_2 x_2(t) + u^2(t), \quad (29a)$$

$$x_2(t+1) = x_1^2(t), \quad (29b)$$

$$y(t) = x_1(t) + e(t). \quad (29c)$$

Using the Ritt's algorithm yields quite lengthy results and below we only present the linear regression equation $\eta_1 = \varphi_1 \theta_1 + \nu_1$ for the first parameter, where

$$\eta_1 = \ddot{u}^2 \ddot{y}^2 - \dot{u}^2 (y^{(3)})^2 - \ddot{y}^2 \dot{y} + (y^{(3)})^2 y,$$

$$\varphi_1 = (y^{(3)})^2 \dot{y} - \dot{y}^3,$$

and

$$\begin{aligned} \nu_1 = & (e^{(3)})^2 e - \ddot{e}^2 \dot{e} - \ddot{e}^2 \ddot{u}^2 + (e^{(3)})^2 \dot{u}^2 - 2e^{(3)} e y^{(3)} \\ & - 2e^{(3)} \dot{u}^2 y^{(3)} + e (y^{(3)})^2 + 2\ddot{e} \dot{e} \dot{y} + 2\ddot{e} \ddot{u}^2 \dot{y} - \dot{e} \dot{y}^2 \\ & - (e^{(3)})^2 y + 2e^{(3)} y^{(3)} y + \ddot{e}^2 \dot{y} - 2\ddot{e} \dot{y} \dot{y} \\ & + \left[\ddot{e}^3 - (e^{(3)})^2 \dot{e} + 2e^{(3)} \dot{e} y^{(3)} - \dot{e} (y^{(3)})^2 \right. \\ & \left. - 3\ddot{e}^2 \dot{y} + 3\ddot{e} \dot{y}^2 + (e^{(3)})^2 \dot{y} - 2e^{(3)} y^{(3)} \dot{y} \right] \theta_1 \end{aligned}$$

We notice that the linear regression becomes quite complicated, but the main complexity lies in the transformed noise. This means that the main problem lies in choosing the appropriate instruments. These equations enable a way to get an initial estimate of the first parameter, which later can be refined using the PEM. \square

6. CONCLUSIONS AND FUTURE WORK

In this paper a discrete-time Ritt's algorithm, similar to the one given in continuous-time, has been presented. It turns out that similar identifiability results as the ones given in Ljung and Glad (1994) can be produced. Also, using these kind of elimination algorithms, in discrete-time, yields the possibility of dealing with noise. Certain aspects of the noise transformation have been discussed from a system identification point of view.

In the future, it would be interesting to examine the possibilities to tailor Ritt's algorithm for system identification purposes. Also, the connection between over-parametrization and Ritt's algorithm, as implicated in Section 5.1, needs further thought. Also, if the resulting linear regression should be used for parameter estimation, how should one choose the instruments to get unbiased estimates?

7. ACKNOWLEDGEMENT

The authors would like to thank the Swedish Research Council and the Vinnova Industry Excellence Center LINKSIC for support.

REFERENCES

- S. Diop. A state elimination procedure for nonlinear systems. In J. Descusse, M. Fliess, A. Isidori, and D. Leborgne, editors, *New trends in Nonlinear Control Theory, Lect. Notes Control Inform. Sci. 122*, pages 190–198. Springer, 1989.
- S. Diop. Elimination in control theory. *Math. Control Signals Systems*, 4:17–32, 1991.
- M. Fliess and S. T. Glad. An algebraic approach to linear and nonlinear control. In H. L. Trentelman and J. C. Willems, editors, *Essays on control: Perspectives in the theory and its applications.*, pages 223–267. Birkhäuser, 1993.
- S. T. Glad. Implementing Ritt’s algorithm of differential algebra. In *IFAC Symposium on Control Systems Design, NOLCOS’92*, pages 610–614, Bordeaux, France, June 1992.
- E.R. Kolchin. *Differential Algebra and Algebraic Groups*. Academic Press, New York, 1973.
- S. Kotsios. An application of Ritt’s remainder algorithm to discrete polynomial control systems. *IMA Journal of Mathematical Control and Information*, 18(18):19–29, 2001.
- L. Ljung. *System Identification, Theory for the User*. Prentice Hall, PTR, 2nd edition, 1999.
- L. Ljung and S. T. Glad. On global identifiability of arbitrary model parameterizations. *Automatica*, 30(2): pp 265–276, February 1994.
- J. F. Ritt. *Differential Algebra*. American Mathematical Society, Providence, R I, 1950.
- A. Seidenberg. An elimination theory for differential algebra. In F. Wolf, J. L. Hodges, and A. Seidenberg, editors, *University of California Publications in Mathematics: New Series*, pages 31–66. University of California Press, Berkeley and Los Angeles, California, 1956.
- T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Springer-Verlag, 1980.

	Avdelning, Institution Division, Department Division of Automatic Control Department of Electrical Engineering	Datum Date 2009-04-03
	Språk Language <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English <input type="checkbox"/> _____	Rapporttyp Report category <input type="checkbox"/> Licentiatavhandling <input type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input checked="" type="checkbox"/> Övrig rapport <input type="checkbox"/> _____
URL för elektronisk version http://www.control.isy.liu.se		LiTH-ISY-R-2891
Titel Identification Aspects of Ritt's Algorithm for Discrete-Time Systems Title		
Författare Christian Lyzell, Torkel Glad, Martin Enqvist, Lennart Ljung Author		
Sammanfattning Abstract <p>In system identification, the challenge of parameter estimation often lies in solving a non-convex optimization problem. In many cases, this implies that it is difficult to guarantee that the global optimum will be found. The tools of differential algebra, for example, Gröbner bases and Ritt's algorithm, have turned out to be quite useful when dealing with certain nonlinear model structures. Some examples of successful applications are the determination of controllability, observability and global identifiability of these model structures. In this paper, difference algebraic techniques, which mimics the differential algebraic methods, will be presented. Besides making it possible to handle discrete-time systems, this opens up the possibility of dealing with noise. It turns out that the classical instrumental variables method plays a role.</p>		
Nyckelord Keywords System identification, Identifiability, Ritt's algorithm.		