INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR. 201 / 23

SECOND WORKSHOP ON MATHEMATICS IN INDUSTRY

(2 - 27 February 1987)

IDENTIFICATION OF LINEAR DYNAMICAL SYSTEMS - III

M. HAZEWINKEL
CWI
P.O.Box 4057
1009 AB Amsterdam
NETHERLANDS

8 Entremets

So far, in the previous two lectures I have (mainly) talked about an identification algorithm based on ideas of Guidorzi [3]. This is very definitely an off-line and non recursive algorithm. I.e. it is not applicable in the following sense. Suppose we have data coming continuously. At time t say we have a (preliminary identification). At time $t+\varepsilon$ a further bit of data comes in. Now of course we could add this to the original data and run the whole procedure again. This is (computer) time consuming and seems inefficient. Also there may not be enough real time to do it. It would be much more efficient (and necessary in fact in certain contexts) to take the original time t identification result, add the new data and use these to update the identification estimate (using only the previous result and the new data and nothing else). If this can be done we speak of a recursive identification procedure, possibly of an on line identification procedure.

There are many ~~silent~~ situations in the technological world today where such a procedure seems the only reasonable one. For example

- Flying an aeroplane: data come in at a furious rate and there is no time for anything but recursive procedures. ~~[illegible]~~ An example from the construction track would be using two helicopters to lift very heavy construction units
- Dealing with systems which change in time where the change in system is slow with respect to the data rate
- Controlling a nuclear reactor

In order to think and reason about this sort of potential algorithm it is useful to know more about the equivalence relation $(A,B,C) \sim (A',B',C')$ iff $\exists\, S \in GL_n(\mathbb{R})$ such that $A' = SAS^{-1}$, $B' = SB$, $C' = CS^{-1}$ (Here $GL_n(\mathbb{R})$ is of course the space of all invertible $n\times n$ matrices). Or more precisely it would be good to know as much as possible about the quotient space

$$L^{co,cr}_{m,n,p}/\sim \;=: L^{co,cr}_{m,n,p}/GL_n(\mathbb{R}) =: M^{co,cr}_{m,n,p}$$

obtained by identifying each equivalence class of systems to one point.

9 The moduli space $M^{co,cr}_{m,n,p}$

Let, as before, $L_{m,n,p} = \{(A,B,C) : A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}\}$ and let $L^{cr}_{m,n,p}$ be the (open) subspace of all such triples such that (A,B) is completely reachable and $L^{cr,co}_{m,n,p}$ the open subspace of all triples which are both co. and cr.

The equivalence relation is given by an action of $GL_n(\mathbb{R})$ on $L_{m,n,p}$, i.e. a (continuous) map $GL_n(\mathbb{R}) \times L_{m,n,p} \to L_{m,n,p}$, $(S,(A,B,C)) \mapsto (A,B,C)^S = (SAS^{-1}, SB, CS^{-1})$ such that $(A,B,C)^{I_n} = (A,B,C)$ and $((A,B,C)^{S})^{T} = (A,B,C)^{TS}$.

A priori taking such a quotient (by a noncompact group) can yield very nasty quotients. However that does not happen here.

(9.1) Theorem. The spaces $M^{cr}_{m,n,p}$ and $M^{co,cr}_{m,n,p}$ are smooth differentiable manifolds of dimension $mn+np$. Moreover the projections π

$$\begin{tikzcd} L^{cr}_{m,n,p} \arrow[d, "\pi"] & \supset & L^{cr,co}_{m,n,p} \arrow[d, "\pi"] \\ M^{cr}_{m,n,p} & \supset & M^{cr,co}_{m,n,p} \end{tikzcd}$$

are smooth principal $GL_n(\mathbb{R})$ fibre bundles (I.e. (among other things) locally the situation looks like

$$\begin{tikzcd} L \arrow[d, "\pi"] & \supset & U \times GL_n(\mathbb{R}) \arrow[d, "\mathrm{proj}_1"] \\ M & \supset & U \end{tikzcd}$$

In fact the situation is even nicer in that the $M^{co,cr}_{m,n,p}$ and $M^{co}_{m,n,p}$, $M^{cr}_{m,n,p}$ are all smooth algebraic varieties defined over $\mathbb{Z}$ and they are fine moduli spaces for a suitable classification problem. This last bit roughly means that there are smooth universal families of systems over them in which each equivalence class has precisely one representant and that to understand how systems change as the parameters vary it suffices to study just this one family.

For all these facts and a discussion of what they mean cf. [7] (appended to these notes)

10 Continuous canonical form.

As remarked several times in these notes already most of the difficulties of identification come from the fact that the input-output data do not determine (A,B,C) uniquely but only up to the equivalence relation $(A,B,C) \sim (A,B,C)^S$ for all $S \in Gl_n(\mathbb{R})$

Many problems would disappear if there were a way of selecting in some good (canonical) way precisely one representant of each equivalence class: ~~By~~ a canonical form. As you know from linear algebra, canonical forms can be very useful (Jordan canonical form)

Then we would have a well defined map

(10.1) (Input output data of Σ) $\mapsto$ (Equivalence class of systems) $\mapsto$ (Canonical representing system (A,B,C))

The algorithm described in the first two lectures is precisely such a map. Now often we will want the result of our identification procedure for other things. Stabilization, (to use) filtering, prediction, model matching, tracking, ... Then it becomes important that the total map (10.1) be continuous, ~~to~~ ~~I~~ ~~be~~ yielding a continuous canonical form on $M^{co,cr}_{m,n,p}$. To be precise:

10.2 Definition. A continuous canonical form on $M^{co,cr}_{m,n,p}$ is a map

$$c: M^{co,cr}_{m,n,p} \to M^{co,cr}_{m,n,p}$$

such that

(i) $c(A,B,C) \sim (A,B,C)$

(ii) $c(A,B,C) = c(A',B',C') \iff (A,B,C) \sim (A',B',C')$

(iii) c is continuous

10.3 Theorem. There exists a continuous canonical form on $M^{co,cr}_{m,n,p}$ ~~iff~~ if and only if $m=1$ or $p=1$

This is not good news. However there are very large subspaces (open and dense) on which canonical forms which are continuous do exist. These are precisely defined by the nice selections (if $p \geq 2$)

Thus the map defined by the identification algorithm of the first two lectures is discontinuous. Not ideal but of course what matters is that it be continuous in the neighborhood of the true system (which we do not know yet) thus picking the right nice selection is important and the Kronecker nice selection (output) may not be the right one. More precisely what looks like the Kronecker nice selection on the basis of our imperfect data may not be the right one.

For more about canonical forms and such and related matters cf. [7, 4, 2]. The last two papers are also appended to these notes

11 Recursive identification schemes

As argued in section 8 above there are many situations in which one would like to have a recursive on-line identification scheme.

Roughly such a scheme works as follows. At time t we have an (preliminary) identification $A(t)$, $B(t)$, $C(t)$ as well as an estimate $\hat{x}(t)$, for the current state of the system. Then we can predict say the next r outputs if this were the true system

$$(11.1)\qquad \begin{aligned} \hat{x}(t+i) &= A\hat{x}(t+i-1) + Bu(t+i-1) \\ \hat{y}(t+i) &= C\hat{x}(t+i) \end{aligned}$$

These predicted outputs can be compared to the actual data as they come in leading to error function, e.g.

$$(11.2)\qquad \sum_{i=t+1}^{t+r} \frac{1}{r} \det(\varepsilon_{ti}\varepsilon_{ti}^T) \qquad \varepsilon_{ti} = \hat{y}(t+i) - y(t+i)$$

or using previous errors as well (and $r = 1$)

$$(11.3)\qquad \sum_{1}^{t+1} \frac{1}{t+1} \det(\varepsilon_i \varepsilon_i^T)$$

(There are here a good many choices one can make involving among other things different schemes of discounting past data)

This defines a function and the idea is now to update the system by moving in the direction of the gradient of this function.

One can for instance think about this as follows. Consider the space

$$(11.4)\qquad M_{1,m,n,p}^{co,cr} = L_{1,m,n,p}^{co,cr} / GL_n(\mathbb{R})$$

of all equivalence classes of quadruples (x, A, B, C) under the equivalence relation

$$(11.5)\qquad (x, A, B, C) \sim (x, A, B, C)^S = (Sx, SAS^{-1}, SB, CS^{-1})$$

where (A, B, C) is supposed to be cr and co. Let there be given a set of input-output data $u(t)$, $y(t)$. Then at every point of

$$(11.6)\qquad M_{1,m,n,p}^{co,cr} \times \mathbb{Z}$$

represented by say $((x, A, B, C), t)$ consider the predictions (say r)

$$(11.7)\qquad \begin{aligned} \hat{x}(t+i) &= Ax(t+i-1) + Bu(t+i-1) & i &= 1, \dots, r \\ \hat{y}(t+i) &= Cx(t+i) & x(t) &= x \end{aligned}$$

and the prediction errors

$$(11.8)\qquad \varepsilon_{ti} = \hat{y}(t+i) - y(t+i)$$

Considering some norm-like function of the ε_{ti} such as $\|\varepsilon_{ti}\|^2$ or the one mentioned above one obtains an error prediction function on the space (11.6) and one is looking for a global minimum of this function. For that one uses something like a steepest descent method which means moving along a gradient. Very similar considerations apply for continuous time but then of course in (11.6) the factor $\mathbb{Z}$ must be replaced by $\mathbb{R}$.

The question now arises whether there are general results guaranteeing the convergence of such schemes to the true system if the data did in fact come from such a system (Consistency results). There are such results "locally", i.e. if one starts out too far from the true system. A much harder question is whether such procedures will converge to a best approximation also if the data do not come from a true system within the class under consideration.

The topology of the space $M^{co,cr}_{1,m,n,p}$ has a role to play here. Indeed

11.9 Theorem (Milnor) Let M be a smooth manifold with a flow on it with one globally attracting equilibrium point. Then M is diffeomorphic to an $\mathbb{R}^n$.

So in particular M is then homotopically trivial. In the continuous time case our procedure defines a flow on $M^{co,cr}_{1,m,n,p} \times \mathbb{R}$ which is homotopically the same as $M^{co,cr}_{1,m,n,p}$ which in turn is homotopically the same as $M^{co,cr}_{m,n,p}$ which if $p \geq 1$ and $m \geq 1$ has nonzero 1-st homotopy group. So in that case such a globally converging flow can not exist.

It remains open whether the data will almost surely define a flow with precisely one sink and all other critical points of source or saddle point type. This depends on the topology of $M^{co,cr}_{1,m,n,p}$ and also on its geometry through the natural embeddings

(11.10)
$$M^{co,cr}_{1,m,n,p} \hookrightarrow \text{Hankel matrix space}$$
$$M^{co,cr}_{m,n,p} \hookrightarrow \text{Hankel matrix space}$$

given by the Hankel matrices associated to (x,A,B,C) and (A,B,C) which are made up out of the matrices and vectors CA^iB and CA^ix, (which are of course all invariants and in fact the only invariants (in the sense that all others are functions of these)).

Incidentally in the above I have glossed over the fact that for a function on a smooth manifold talking about the gradient vector field of a function requires having a Riemannian metric defined on it. This can e.g. be done by using the embeddings (11.10) and taking the induced metric coming from a metric on Hankel space $\{(H_0, H_1, \ldots)\}$, e.g. the Euclidean one or possibly $\Sigma\, tr(H_k H_k^T)$.

Such recursive identification algorithms have been proposed by Overbeek-Ljung [5] and Hanzon [6], complete with local convergence / consistency results. The just named one has also been implemented as a computer program and seems to work well.

In any case recursive identification procedures involve "walking around on the manifolds $M^{cr,co}_{m,n,p}$ and closely related manifolds such as $M^{cr,co}_{1,m,n,p}$ and defining suitable flows on it. There are a number of questions which now come up which seem natural.

(i) Convergence analysis. Almost sure convergence. It would be a major step forward to be able to prove a convergence result for the given coordinate patches U_α corresponding to a nice (output) selection. (Because the U_α are open and dense)

(ii) Are there holes in $M^{co,cr}_{m,n,p}$. For a discussion of this matter and results cf [17] (appended to these notes)

(iii) Given a flow, i.e. a differential equation on a noncompact manifold, are there ways to control the errors induced by coordinate changes. An interesting subquestion here is whether $M^{co,cr}_{m,n,p} \subset$ Hankel space is asymptotically flat.

(iv) As remarked before much of our trouble comes from the fact that we have to work on a quotient space of $L^{co,cr}_{m,n,p}$ on which there is no global coordinate chart. On $L^{co,cr}_{m,n,p}$ there is no trouble in defining a suitable "identification flow" (degeneracy of error prediction function)

But it may be possible to find intermediate quotients

$$\begin{tikzcd} & L^{(0,c)}_{m,n,p} \arrow[dl] \arrow[d] \\ M \arrow[r] & \Pi^{(0,c)}_{m,n,p} \end{tikzcd}$$

such that e.g. $M \to \Pi^{(0,c)}_{m,n,p}$ has desirable fibres and such that M has a global coordinate chart. Then on M a suitable flow would be induced and that would work just as well.

An intermediate quotient with compact fibres would also be useful in connection with (iii) above (and global coordinates)

In all of the above I have neglected or ignored the matter of noise in the observations and possibly additional input noise fed in through a matrix $B_2 w(t)$ where $w(t)$ is a series of identically distributed Gaussian random variables.

There are various ways to take this into account and the presence of noise does not seem to complicate matters essentially much.

12. The nonlinear filtering approach to identification

~~Here~~ The idea is to look at a linear unknown system

$$(12.1) \qquad \begin{aligned} dx &= Ax\,dt + B_2\,dw + B_1\,du \\ dy &= Cx\,dt + dv \end{aligned}$$

as a nonlinear system with state vector (x, A, B_1, B_2, C) which state has to be estimated (filtered) on the basis of the observations $y(s)$, $0 \le s \le t$ for each t. Again we shall be interested in recursive ways of doing this.

~~For~~ One approach to this problem originally proposed by Brockett, Clark and Mitter which I consider very promising, cf. [8] which is appended to these notes.

References

[1] M. Hazewinkel, R.E. Kalman, Invariants, canonical forms and moduli for linear, constant finite dimensional dynamical systems. In: G. Marchesini, S.K. Mitter (eds), Proc. of a Conf. on algebraic System Theory, Udine 1975, Springer 1976, 48-60

[2] M. Hazewinkel, On the (internal) symmetry groups of linear dynamical systems. In: P. Kramer, M. Dal-Cin (eds), Groups, systems and many-body physics, Vieweg 1979, 362-404

[3] R. Guidorzi, Complete sets of independent invariants and canonical forms for linear systems identification, Research Report Jan 1978, Inst. di Automatica, Univ. di Bologna

[4] M. Hazewinkel. Parametrization problems for spaces of linear input-output systems. In: H. Nennzert (ed), The Road-Vehicle system, Teubner, 1986

[5] A.J.M. van Overbeek, L. Ljung, On-line structure selection for multivariable state-space models, Report LiTH-ISY-I-0393, Linköping Univ. and Automatica 1982

[6] B. Hanzon, Identifiability, recursive identification and spaces of linear dynamical systems. Thesis, Erasmus Univ. Rotterdam, 1986

[7] M. Hazewinkel, (Fine) moduli (spaces) for linear systems: what are they and what are they good for. In: C.I. Byrnes, C.F. Martin (eds), Geometrical methods for the theory of linear systems, Reidel, 1980, 125-193

[8] M. Hazewinkel, Lie algebraic methods for filtering and identification. To appear in Proc. 1-st world congress of the Bernoulli society, Tashkent, 1986, VNU Science Press, and in Proc. VIII-th IAMP Congress, Marseille, 1986, World Scientific.

[9] R.K. Mehra, Identification in control and econometrics: similarities and differences, Ann. Econ. and Social Measurement 3:1 (1974), 21-47

[10] R.K. Mehra, Identification of linear dynamical systems II. In: M. Hazewinkel, A.H.G. Rinnooy Kan (eds), Ten lectures on the interface Economics, Econometrics, Mathematics, Reidel Publ. Co., 1982

[11] M. Hazewinkel, On families of linear systems: degeneration phenomena. In: C.I. Byrnes, C.F. Martin (eds), Algebraic and geometric methods in linear system theory, AMS, 1980