International Journal of Theoretical Physics, Vol. 34, No. 8, 1995

Logics in Realization and Modeling from Data

Erik I. Verriest[1]

Received October 22, 1994

System *realization* is the construction of a state-space model given input–output data of a system. One approach, briefly summarized here, is the subspace method. In the deterministic realization problem, the data are used in a *linear* fashion, whereas the stochastic realization problem uses *quadratic* forms in the data. This dichotomy is related to the basic assumptions of repeatability or nonrepeatability of the input–output experiments performed on the system. In particular, the logic of the system is constructed, closely following the axiomatic foundations of physics. It is shown that this logic is Boolean in the deterministic and quantal in the stochastic case. The system dynamics is obtained from the data-induced measures one can define on the lattice.

1. INTRODUCTION

Many models of reality for use in engineering and science are constructed from established natural laws. The immediate applications are to (i) explain the past, (ii) predict the future, and (iii) control. For many systems, however, the cause and effect relationship is either too complicated or too obscure to yield a practical representation. This motivates the need for empirical methods of finding black-box models.

The usefulness of any model is determined not only by its predictive power or accuracy, but also by its mathematical tractability. Because of this, linear time-invariant (LTI) systems have been intensively studied. Modeling the system dynamics from such classical representations as its transfer function (frequency response) or its impulse response is a relatively simple problem in theory (Kailath, 1980). However, these are idealized representations, and are often unavailable in practice. One therefore seeks a more pragmatic approach. A promising strategy in modeling of LTI systems from input–output data is the subspace method. A close inspection of the algorithm reveals that

[1]Georgia Tech Lorraine, F-57070 Metz, France.

0020-7748/95/0800-1775$07.50/0

in the deterministic case, these data are used in a *linear* fashion, while for the stochastic case, the data are used *quadratically.* In this paper we shall explain this fundamental difference by relating it to the *logic* of the system and the measures defined on it.

In Section 2, we pose the realization problem in detail for the class of linear time-invariant systems. The subspace approach toward the solutions for the deterministic and stochastic realization problems is summarized in Section 3. In Section 4, an interpretation based on the underlying logic is given in both problems. Deterministic realization relates to stochastic realization as classical logic to quantum logic. This indicates that quantum structures are also prevalent in other well-established disciplines such as signal processing and time series analysis.

2. REALIZATION PROBLEM

For simplicity of the exposure, it is assumed that a system Σ is accessed via a scalar-valued input and produces a scalar output, and that the interactions (transitions, measurements) occur at discrete instants only. All these restrictions are removable in the complete theory.

Assume that the only concrete knowledge one has about the system is its consecutive input and output data. It is not necessary that one can actually *set* (control) the inputs, but it is assumed that whatever goes in can be *measured.* Assume further that contextual information suggests time (as measured by counting the transitions) invariance.

The goal is to describe the input and output data by a simpler rule, using a compressed set of parameters and expressed in a mathematically tractable form. The class of linear time-invariant state-space model is chosen, i.e., the form

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = Cx_k + Du_k + v_k \end{cases} \tag{1}$$

Time invariance manifests itself in the constancy of the matrices of equation (1). The $w(t)$ and $v(t)$ are the 'residuals' due to imprecision in the measurements, or incorrectness of the model class assumption, e.g., if environmental interactions with the system are purposely neglected. The free parameters in the model are the model order n as well as A, B, C, and D.

The realization problem consists in retrieving these free parameters from the sequences $\{u_k\}$, $\{y_k\}$ in the deterministic case, and augmenting them with the stochastic parameters (covariances) of the environmental interactions in the stochastic case.

3. SUBSPACE METHOD

One considers in the *deterministic* problem, the sequence of input–output pairs

$$d_k = \begin{bmatrix} u_k \\ y_k \end{bmatrix}$$

and forms its associated (block) Hankel matrix

$$\begin{bmatrix} d_1 & d_2 & \cdots & d_T \\ \vdots & \vdots & & \vdots \\ d_j & d_{j+1} & \cdots & d_{2T-1} \end{bmatrix} = P \begin{bmatrix} u_1 & u_2 & \cdots & u_T \\ \vdots & \vdots & & \vdots \\ u_T & u_{T+1} & \cdots & u_{2T-1} \\ -- & -- & -- & -- \\ y_1 & y_2 & \cdots & y_T \\ \vdots & \vdots & & \vdots \\ y_T & y_{T+1} & \cdots & y_{2T-1} \end{bmatrix} \tag{2}$$

where P is a (row) permutation matrix. The relevance of the above Hankel matrix is easily explained as follows: From the underlying model (1) the generated (nominal) *data sequence* of length T is

$$\begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+T-1} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{j-1} \end{bmatrix} x_k + \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \ddots & \ddots & \\ CA^{T-2}B & \cdots & CB & D \end{bmatrix} \begin{bmatrix} u_k \\ u_{k+1} \\ \vdots \\ u_{k+T-1} \end{bmatrix} \tag{3}$$

or $\mathcal{Y} = \mathcal{O}x + \mathcal{T}\mathcal{U}$. Let us now assume that many (identical) systems are placed in parallel, each with its own input sequence $\mathcal{U}^{(i)}$ and generating the corresponding output sequence $\mathcal{Y}^{(i)}$ for $i = 1, \ldots, N$. Putting likewise the individual relations (3) for each i in parallel, one obtains in matrix form

$$[\mathcal{Y}^1, \ldots, \mathcal{Y}^N] = \mathcal{O}[x^{(1)}, \ldots, x^{(N)}] + \mathcal{T}[\mathcal{U}^1, \ldots, \mathcal{U}^N] \tag{4}$$

or $\overline{\mathcal{Y}} = \mathcal{O}\overline{X} + \mathcal{T}\overline{\mathcal{U}}$. To work with a statistically representative sample, N must be larger than the number of consecutive data points (T) in the sequence. In practice, an ensemble of systems may not be available. An artificial ensemble of input/output data parallel experiments can be made from one long data sequence via *time shifting*. If $\mathcal{Y}^{(1)}$ is the sequence of T consecutive output samples starting at k, then $\mathcal{Y}^{(2)}$ is the sequence of T consecutive samples starting with y_{k+1}. Regardless of how the ensemble data is generated, the data are organized in a so-called *data matrix*,

$$H = \begin{bmatrix} \overline{\mathcal{Y}} \\ \overline{\mathcal{U}} \end{bmatrix} = \begin{bmatrix} \mathcal{O}\overline{X} + \mathcal{T}\overline{\mathcal{U}} \\ \overline{\mathcal{U}} \end{bmatrix} \tag{5}$$

It was shown that under some conditions (observability and sufficient excitation of the modes by the inputs) (De Moor *et al.*, 1988), the property

$$\text{rank}(H) = \text{rank}(\overline{\mathcal{U}}) + n \tag{6}$$

holds, where n is the order of the underlying system (1) that generated the data. What is now involved in the subspace realization algorithm? First, one determines the state vectors (a basis for the state space). Numerically, singular value decomposition of the data matrix is used. Then, consistently with the data and the derived state vectors, a quadruple (A, B, C, D) is obtained as the solution to a *total least squares problem* (Van Huffel and Vandewalle, 1991). This two-part solution is now briefly described.

State Space Basis

One forms *two* consecutive data matrices

$$H_1 = \begin{bmatrix} d_k^{(1)} & \cdots & d_k^{(N)} \\ \vdots & & \vdots \\ d_{k+T-1}^{(1)} & \cdots & d_{k+T-1}^{(N)} \end{bmatrix} = P \begin{bmatrix} \overline{\mathcal{U}_1} \\ -- \\ \overline{\mathcal{Y}_1} \end{bmatrix} \tag{7}$$

$$H_2 = \begin{bmatrix} d_{k+T}^{(1)} & \cdots & d_{k+T}^{(N)} \\ \vdots & & \vdots \\ d_{k+2T-1}^{(1)} & \cdots & d_{k+2T-1}^{(N)} \end{bmatrix} = P \begin{bmatrix} \overline{\mathcal{U}_2} \\ -- \\ \overline{\mathcal{Y}_2} \end{bmatrix} \tag{8}$$

Applying the rank property (6) to H_1, H_2, and $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$, the system order n can be inferred, and one can derive, invoking Grassmann's dimension theorem, that the state space is the intersection $X_2 = \text{Span}_{\text{row}}(H_1) \cap \text{Span}_{\text{row}}(H_2)$. Although in general the rowspaces may not intersect due to finite precision, a likely intersection may be found using the powerful singular value decomposition (SVD) technique. Any basis for this rowspace is then a realization of X_2.

System Matrices

With any choice of a basis for X_2, a realization quadruple corresponds via

$$\begin{bmatrix} x_{k+T+1}^{(1)} & \cdots & x_{k+T+1}^{(N)} \\ y_{k+T}^{(1)} & \cdots & y_{k+T}^{(N)} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_{k+T}^{(1)} & \cdots & x_{k+T}^{(N)} \\ u_{k+T}^{(1)} & \cdots & u_{k+T}^{(N)} \end{bmatrix} \tag{9}$$

This is an overdetermined set of equations (since N is large), and is solved in the least squares sense, or better, the total least squares sense (since the left- and right-hand-side matrices have similar accuracies, and are therefore equally perturbable). The residuals are the $\{w_k\}$ and $\{v_k\}$ noise terms in the original model. It was shown in Moonen *et al.* (1989) that the explicit

computation of X_2 may be bypassed in practice. Indeed the known matrices on the left- and right-hand sides of equation (9) can be replaced by, respectively, H_L^{der} and H_R^{der}, which are derived via the SVD on the original data matrices.

The important conclusion is that the observed data are used in a *linear* fashion in the construction of a state space model.

Now consider the *stochastic* realization problem. Conceptually, the *measured* input and output data are different from the *actual* input and output in the system. The difference is usually ascribed to noise, and the methodology is known as the *error-in-variables* configuration. In a *white noise* model, these perturbations are assumed to have zero mean and be timewise uncorrelated. The covariance information of these input and output fluctuations is summarized in the covariance matrix

$$\text{Cov}\begin{pmatrix} w_k \\ v_k \end{pmatrix} = \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \geq 0 \tag{10}$$

Essentially, the free system parameters (A, B, C, D) representing the actual dynamics in such a stochastic model are augmented with the stochastic parameters (Q, R, S).

Using this probabilistic model, a first step in the solution to this stochastic realization is to obtain the correlations of the fluctuations

$$\Lambda_l = \mathbf{E}\begin{pmatrix} \tilde{u}_k \\ \tilde{y}_k \end{pmatrix}(\tilde{u}_{k+l}^*, \tilde{y}_{k+l}^*) \tag{11}$$

and form the Hankel matrix H_{stoch} of this *correlation* sequence. Letting the above-described subspace algorithm loose on this data matrix, one finds a quadruple (A, B, C, D). The stochastic parameters are obtained by solving for P, Q, R, and R in a set of linear matrix equalities with the positive semidefiniteness constraint of (10) and $P > 0$. When working with empirical data, the exact correlations are not available, and one has to work with the *sample covariances*. This means that the algorithm uses the original data in a *quadratic* fashion.

One sees a fundamental difference in the usage of the experimental data. It is a matter of using the observed or measured variables *linearly versus quadratically*. Also, if the experimenter only has the experimental data available, then what prompts the choice *deterministic versus stochastic* model?

We will turn to the *logic* of the system to partially get an answer to these questions. Why the data are used linearly or quadratically can be resolved this way, but one will always have to make a basic assumption as to whether the experiment is repeatable or not at the onset. This of course is where inductive reasoning enters modeling.

4. INTERPRETATION IN QS

In order to try to explain the linear-quadratic dichotomy, we give an axiomatic setup for systems (Verriest, 1992), which closely parallels the modern axiomatic foundations of physics (Beltrametti and Cassinelli, 1981; Piron, 1990).

The *observable* quantities (attributes) of a system (i.e., its input and output) or environmental variables are considered primordial. Observable means here that the experimenter can extract information from these. This process is called a *measurement.* It is assumed that physical measurements can resolve whether or not the value assumed by the variable lies in some preset (by the experimenter/observer) interval B (or a finite set if the observable assumes discrete values), and that there is *no* reaction from the measurement to the variable. Once a variable has been measured, the information is not erased, but the stage is set for perhaps another measurement. Measurements are "yes–no" experiments, but at a lower level than considered by Macky. We shall call the study of the relations among measurements on variables the *kinematics.* All propositions that can be made about a set of past measurements form a Boolean lattice.

By a *filter* we understand a device that 'accepts' a system if the measurement falls within a preset interval. Acceptance is with respect to consequent use, as member of a particular set. The filter $u_k \in B$ describes an ensemble of systems, all having the *property* that at time k the input value belonged to B.

Now we pass from the kinematics of variables to the *dynamics* of a system. A *physical preparation* of a system relates to the notion of a filter. Only these copies of an ensemble of systems are accepted (for further consideration) that passed a fixed filter (on input and/or output). In a *control* point of view, e.g., if one has access (externally, via u) to the system, only copies of the ensemble are considered that have this same input (as can be measured and accepted by a filter tuned to this input). From a 'selection' point of view, a filter characterized by a finite sequence of data $(d_1, \ldots, d_T)$, strictly speaking, a sequence of data *intervals*, is chosen. In this passive way, only the systems passing this filter are retained for further scrutiny. The notion of preparation allows the transition from the kinematics to the dynamics in the theory. Let Π denote the set of all preparations. Next we consider the *propositional system.*

A fundamental entity is an *s-question*, corresponding to a statement about the I/O attributes of the system. Just like questions, the s-questions have answers that are either 'yes' or 'no' when pertaining to one concrete instance of the system. The s-question is *true* if the answer is 'yes' in all instances of the ensemble of systems that passed a given filter. If 'no' is

possible, the s-question is not true. The s-questions test *properties of the system*, rather than *values of a variable*. A property is *actual* if the corresponding s-question is true, in the other case it is *potential*. Let $\mathscr{E}$ denote the set of all s-questions that may be asked about the system. All knowledge of the system stems then from knowledge of the map $\mathscr{E} \times \Pi \to \{0, 1\}$. However, the sets $\mathscr{E}$ and Π are too large. Denote by $\pi(e)$ the answer (**1** or **0** for 'actual' or 'potential') to question e if the system was prepared by π. There exist equivalences $\overset{e}{\sim}$ and $\overset{p}{\sim}$ defined by

$$e_1, e_2 \in \mathscr{E}: e_1 \overset{e}{\sim} e_2 \Leftrightarrow \forall \pi \in \Pi: \pi(e_1) = \pi(e_2) \tag{12}$$

$$\pi_1, \pi_2 \in \Pi: \pi_1 \overset{p}{\sim} \pi_2 \Leftrightarrow \forall e \in \mathscr{E}: \pi_1(e) = \pi_2(e) \tag{13}$$

Equivalences lead to reductions to quotient sets, and we call $\mathscr{L} = \mathscr{E}/\overset{e}{\sim}$ the set of all *propositions*, and $\mathscr{S} = \Pi/\overset{p}{\sim}$ the set of all *states*. The set $\mathscr{L}$ is endowed with a lattice structure. The *logic* of the system is a *complete orthocomplemented lattice*. A typical proposition is $\pi \Rightarrow (u^+, y^+)$: If prepared by π, the (next) input u^+ generates output y^+.

The main implication is that for a *repeatable* experiment (deterministic system), there exists a subensemble of systems whose members give identical input–output behavior. This means that we can characterize such a subensemble by a *filter* $(d_1, \ldots, d_T)$ for some T and data $\{d_k\}$. In other words, $(d_1, \ldots, d_T)$ and u_{T+1} completely determine y_{T+1}. The string $(d_1, \ldots, d_T)$ has an ontological meaning, and the property (u^+, y^+) is called *classical*. Cartan's isomorphism shows that $\mathscr{L}$ is a Boolean lattice (Piron, 1990).

For a *nonrepeatable* experiment (stochastic system), there does not exist a finite T such that the filter $(d_1, \ldots, d_T)$ specifies y_{T+1} uniquely, given u_{T+1}. The properties $\pi \to (u^+, y^+)$ are not classical in the sense of Piron. Here one could *deny reality* to all $(d_1, \ldots, d_T)$ since it has no meaning as system state, apart from conditioning (i.e., the filter itself in the kinematics). The only statements that can be made are statements about relative frequencies, which in fact require infinite data, and are therefore again not physical. It follows the newer thought to conclude that there are no *classical* properties, except for the **0** and **1**.

Consider again the ensemble (set of instances of the same system), and let it be prepared by the filter $\pi = (d_1, \ldots, d_T)$. Due to measurement inaccuracies, the property $\pi \Rightarrow (u^+, y^+)$ may be true for some of the instances and not for others. In order to measure our "belief" in a model (based on the total available data set), we shall put a *measure* on the proposition lattice. It allows the evaluation of our subjective uncertainty regarding the dynamics of the system, since the proposition $\pi = (d_1, d_2, \ldots, d_T) \Rightarrow (u^+, y^+)$ is in fact a statement regarding the (potential) state transition u^+: $\pi \to \pi' = (d_2,$

$d_3, \ldots, d_T, d_{T+1}$), where $d_{T+1} = (u^+, y^+)$. These state transition propositions are represented as filters $(d_1, d_2, \ldots, d_T, d_{T+1}) \in D^{T+1}$, where $\forall i, d_i \in D = \mathbb{R}^\delta$.

Implications

For a *single* data sequence $(d_1, d_2, \ldots, d_L)$, where L is finite, consider the *filtration*, i.e., a sequence of successively refining filters (= potential states)

$$d_1, \quad \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \quad \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}, \ldots \tag{14}$$

and create the 'artificial' ensembles fabricated by time shifting subsequences of a single sequence.

If a T exists such that the potential state $(d_1, \ldots, d_T)$ characterizes the I/O behavior completely, then the logic is Boolean, and can be represented by subsets of a set (embedded in $\mathbb{R}^{\delta T}$, where $\delta = \dim d$). The measures on this space (of subsets) are Kolmogorovian. With a data matrix

$$X = [x_1, x_2, \ldots, x_N], \qquad x_i = \begin{bmatrix} d_i \\ \vdots \\ d_{i+T-1} \end{bmatrix} \tag{15}$$

the data-induced measures are of the *counting* form

$$\xi_X(S) = \frac{1}{N} \sum_i \xi_{x_i}(S); \qquad \xi_{x_i}(S) = \chi_S(x_i) = \begin{cases} 1 & \text{if } x_i \in S \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

The dynamic model u^+: $\pi \to \pi^+$ is determined by selecting for each state π the sub*set* of D^{T+1} in a model class which has the largest measure (the most points). If one chose the LTI model class, a suitable class of sets is the set of wedges bounded by two hyperplanes. The angle between the wedges relates to the *tolerated misfit* in the model.

If there does not exist a T such that a potential state (filter) $(d_1, \ldots, d_T)$ characterizes the input–output behavior completely (i.e., the stochastic case), then the logical **1** and **0** are the only classical properties. The corresponding lattice is then a *purely quantum lattice*. The logic is represented by $\mathcal{P}roj(\mathcal{H})$, the set of closed subspaces of a Hilbert space $\mathcal{H}$. The positive-valued measures are the *Gleason measures:* Consider a single data vector $x \in \mathbb{R}^{\delta T}$. The data-induced vector measure (orthogonally scattered measure) is

$$\xi_x\colon \mathcal{P}roj(\mathbb{R}^{\delta T}) \to \mathbb{R}^{\delta T}\colon \xi_x(S) = P^S x \tag{17}$$

where P^S is the projector onto the subspace S. The induced positive-valued measure is then

$$\mu_x(S) = \|\xi_x(S)\|^2 = \text{Tr } P^S xx' \tag{18}$$

For a *data matrix* (many parallel experiments), let

$$X = [x_1, x_2, \ldots, x_N] = \sum_i x_i \otimes e_i \tag{19}$$

One gets a positive-valued measure by a coherent superposition:

$$\xi_X(S) = \sum_i \xi_{x_i}(S) \otimes e_i \tag{20}$$

$$\mu_X(S) = \|\xi_X(S)\|^2 = \text{Tr } P^S \sum_i x_i x_i' = \text{Tr } P^S XX' = \text{Tr } P^S T_X \tag{21}$$

Note that T_X is the (unnormalized) density operator, and it is quadratic in the data.

The introduction of the orthogonally scattered Gleason (OSG) measures is dictated by the necessity of considering a quantum logic for a nonrepeatable experiment. In turn, nonrepeatability and therefore lack of exact predictability is a fundamental attribute of what is usually referred to as stochasticity. It is therefore perhaps not surprising that other standard methods in time series analysis can also be related to quantum structures and concepts. In Verriest and Finkelstein (1991) principal component analysis (also known as Karhunen–Loève) and canonical correlation analysis were related to the maximization, under various constraints, of the RV coefficient of multivariate statistics, which is readily interpreted in terms of quantum structures and data measures.

Let a data matrix X be given and partitioned into 'past' and 'future':

$$X = \begin{bmatrix} X_{\text{past}} \\ X_{\text{future}} \end{bmatrix}$$

Letting $T_X = XX'$ be the unnormalized density matrix, one finds for P^p and P^f, respectively the projectors onto 'past' and 'future':

$$\text{RV('past','future')} = \frac{\text{Tr } T_{12} T_{21}}{(\text{Tr } T_{11}^2 \text{ Tr } T_{22}^2)^{1/2}} \tag{22}$$

$$= \frac{\|P^p T_X P^f\|^2}{\|P^p T_X\| \cdot \|P^f T_X\|} \tag{23}$$

$$= \frac{\langle \xi_X(P^p), \xi_X(P^f) \rangle^2}{\|\xi_X(P^p)\|^2 \|\xi_X(P^f)\|^2} \tag{24}$$

$$= \rho(\xi_X(P^p), \xi_X(P^f))^2 \tag{25}$$

The RV coefficient of past and future is the square of the correlation coefficient between the data-induced OSG measures on the past and the future.

REFERENCES

Beltrametti, E. G., and Cassinelli, G. (1981). *The Logic of Quantum Mechanics,* Addison-Wesley, Reading, Massachusetts.

De Moor, B., Moonen, M., Vandenberghe, L., and Vandewalle, J. (1988). In *Analysis and Optimization of Systems,* A. Bensoussan and J. Lions, eds., Springer-Verlag, Berlin.

Kailath, T. (1980). *Linear Systems,* Prentice-Hall, Englewood Cliffs, New Jersey.

Moonen, M., De Moor, B., Vandenberghe, L., and Vandewalle, J. (1989). On- and off-line identification of linear state-space models, *International Journal of Control,* **49**(1), 219–232.

Piron, C. (1990). *Mécanique quantique–Bases et applications,* Presses polytechniques et universitaires romandes, Lausanne.

Van Huffel, S., and Vandewalle, J. (1991). *The Total Least Squares Problem: Computation Aspects and Analysis*, SIAM.

Verriest, E. I. (1992). Logic, geometry and algebra in modeling, in *Proceedings of the 2nd IFAC Workshop on Algebraic-Geometric Methods in System Theory,* Prague.

Verriest, E. I., and Finkelstein, D. (1991). Unified theory of model reduction via Gleason measures, *International Journal of Theoretical Physics,* **30**, 1703–1714.