

System Identification by Dynamic Factor Models¹

C. Heij²

W. Scherrer³

M. Deistler³

Abstract

This paper is concerned with linear dynamic factor models. In such models the observed process is decomposed into a structured part called the latent process, and a remainder that is called noise. The observed variables are treated in a symmetric way, so that no distinction between inputs and outputs is required. This motivates the condition that also the prior assumptions on the noise are symmetric in nature. We investigate the relation between optimal models and the spectrum of the observed process. This concerns in particular properties of continuity and consistency. Several possible noise specifications and measures of fit are considered.

1 Introduction

In this paper we are concerned with the identification of linear systems. The most commonly used models in system identification are ARMA and ARMAX models, we refer to [15], [3] and [11]. An ARMA model is symmetric and non-open, in the sense that all observed variables are treated in a symmetric way and that they are completely described by the model. On the other side, ARMAX models are non-symmetric and open, as a distinction is made between inputs and outputs and the noise is added to the outputs, and the inputs are not modelled.

We will consider linear factor models where the noise model is symmetric and where we have a deterministic, symmetric and open system model. In a sense these models combine the symmetry which is inherent in, for example, ARMA models, with the flexibility of models that leave certain process aspects unexplained, as for example in input-output models.

¹Support by the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" Projekt P11213-MAT, and by the EC-SCIENCE "System Identification" Project SC1*-CT92-0779 is gratefully acknowledged.

²Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, phone +31-10-4081269, fax +31-10-4527746.

³Institut für Ökonometrie, Operations Research und Systemtheorie, Technische Universität Wien, Argentinierstrasse 8, A-1040 Wien, Austria, phone +43-1-58801-4405, fax +43-1-5054524.

Of course, the classical ARMA and ARMAX models are appropriate in a great number of cases. For instance, if we are interested in predicting the outputs from the inputs then the ARMAX setting is appropriate. On the other hand, there are also situations where this approach can not be justified and may lead to prejudiced results.

- A prediction based error model is not appropriate, for example, if we are interested in the 'true' underlying system and there is noise on the inputs and the outputs.
- There may be uncertainty about the number of system equations or about the classification of the system variables into inputs and outputs. In this case we have to perform a more symmetric way of system modelling, which in turn demands a symmetric noise model.
- In multivariate time series analysis one is confronted with the so-called curse of dimensionality. One method of reducing the dimension of the parameter space for the system model is dynamic factor analysis. This is an essential aspect of the approach described here.

Factor models have been used in statistics, psychometrics and econometrics for a long time, see [8], [1], [9]. The theory is most well-developed for the case of static models. Most applications are also reported within this framework, although there are also contributions on the identification of dynamic factor models, see [10], [7], [4]. Within the area of systems and control there is recently an increasing interest in symmetric modelling. We mention the introduction of the behavioural approach in systems theory in [19], [21], the attention for the Frisch problem, see [16], [18], [2], and low-noise modelling as proposed in [14]. Most contributions on factor models in this area deal either with the mathematical structure of dynamic models or with data modelling by means of static models. In an, in a certain sense, nonparametric framework results on the identification of dynamic factor models within a stochastic setting have been presented in [5], [6]. Procedures for symmetric time series modelling within a deterministic behavioural framework have been proposed in [20], [12], and [17].

In this paper we try to integrate the above two frameworks, i.e., stochastic factor models and deterministic behavioural modelling. The model class consists of stochastic dynamic factor models where the latent process satisfies deterministic behavioural laws. This means that stochastic structure is added to the deterministic behavioural framework, which provides additional tools of analysis. On the other hand, our approach allows for an analysis of dynamic factor models in terms of finite dimensional systems, as opposed to the nonparametric results that were previously obtained.

We now give an outline of the topics treated in this paper. A dynamic factor model is of the form

$$w = \hat{w} + \tilde{w} \quad (1)$$

where w is the observed process, \hat{w} is an (in general unobserved) latent process satisfying exact linear dynamic equations, and \tilde{w} is the noise process. The restrictions on \hat{w} can be expressed in terms of linear deterministic system behaviours as introduced in [19], [21]. The noise process represents the error resulting from the approximation of the observations w by the latent process \hat{w} .

The central question considered in this paper is how to obtain the restrictions satisfied by the latent process from the observations. Without imposing further conditions, no solutions can be excluded from the knowledge of the observed process alone. This means that we have to impose additional assumptions on the noise structure in order to make meaningful statements about the underlying system. In particular we consider optimal models, which under a given bound on the complexity of the linear restrictions, have the smallest possible error.

This paper has the following structure. In Section 2 we define the dynamic factor model. For this purpose we review the behavioural approach in linear system theory. In addition we introduce optimal models, which minimize the noise under restrictions on the complexity of the latent process. Section 3 is concerned with consistency of a certain estimator of the set of optimal models and with the underlying topological problems. Due to restrictions of space we can only give a flavour of the problem in the sense that only a few results are presented and the proofs are omitted. For further information the reader is referred to [13].

2 Dynamic Factor Models

2.1 Linear Systems

For the formulation of dynamic factor models it is convenient to use the behavioural approach as developed

by Willems in [19], [21]. Since this approach may be not well-known to the reader, we discuss in this section those aspects that are relevant for our purposes. Readers with an interest for further details and proofs are referred to [19], [21].

In this subsection $\hat{w} : \mathbf{Z} \rightarrow \mathbf{R}^q$ denotes a trajectory rather than a process, that is, it is a q -variate time series observed in discrete time. The behaviour of a deterministic system is defined as the set of all trajectories \hat{w} that may arise within the restrictions imposed by the system. So a behaviour is a subset \mathcal{B} of $(\mathbf{R}^q)^{\mathbf{Z}}$. Of special interest are behaviours that are linear, time invariant, and complete. This means that $\mathcal{B} \subset (\mathbf{R}^q)^{\mathbf{Z}}$ is a linear subspace that is invariant under the shift operator σ , defined by $(\sigma \hat{w})(t) := \hat{w}(t + 1)$, and that the behaviour is in addition closed in the topology of pointwise convergence. The last condition means that for a sequence $\hat{w}_n \in \mathcal{B}$ which converges pointwise (in \mathbf{R}^q) to $\hat{w}_0 \in (\mathbf{R}^q)^{\mathbf{Z}}$ there holds that also $\hat{w}_0 \in \mathcal{B}$. These conditions imply that the behaviour corresponds to a linear, time invariant, finite dimensional system. In the sequel we will simply use the term linear system to refer to a linear, time invariant, complete behaviour $\mathcal{B} \subset (\mathbf{R}^q)^{\mathbf{Z}}$.

Linear systems can be represented in several ways. Here we discuss representations in terms of polynomial equations and state space models with driving variables.

Every linear system can be represented in polynomial form, as the solution set of the polynomial equations

$$R(\sigma) \hat{w} = 0 \quad (2)$$

Here R is a polynomial matrix in the shift σ . The representation of a given system by a polynomial matrix is highly non-unique.

An alternative representation is in terms of state space models with driving variables. Every linear system can be represented as

$$\sigma x = Ax + Bv, \quad \hat{w} = Cx + Dv \quad (3)$$

Here v is an m -dimensional auxiliary vector of unrestricted driving variables and x is an n -dimensional vector of state variables. In contrast with the usual input-state-output model, here all the external variables are described as outputs of a system driven by forces which need not have any external meaning. For a given system this kind of representation is highly non-unique. A representation (3) is called minimal, if for any representation with m' driving variables and n' states either $m < m'$ or $m = m'$ and $n \leq n'$ holds. The minimal pair (m, n) is used as a measure of the complexity of the system \mathcal{B} .

Until now no assumptions were made concerning the controllability of systems. A system \mathcal{B} is called controllable if every future in \mathcal{B} is attainable from every past in \mathcal{B} , that is, if for every $\hat{w}_1, \hat{w}_2 \in \mathcal{B}$ there exist $\hat{w} \in \mathcal{B}$ and $h \geq 0$ such that $\hat{w}(t) = \hat{w}_1(t)$ for $t < 0$ and $\hat{w}(t) = \hat{w}_2(t)$ for $t \geq h$. In terms of the kernel representations (2) this means that $R(z)$ has constant rank over $z \in \mathbf{C}$. There is a close connection between the notion of controllability as defined before and the usual notion in terms of state space models, because minimal state models (3) of controllable systems \mathcal{B} are characterized by the property that (A, B) is a controllable pair and (A, C) an observable pair.

For controllable systems the transferfunction

$$P_{\mathcal{B}} = I - R^*(\sigma)(R(\sigma)R^*(\sigma))^{-1}R(\sigma) \quad (4)$$

is the orthogonal projection onto the set of square summable time series in the behaviour \mathcal{B} . Here $R^*(\sigma) = R'(\sigma^{-1})$ is the adjoint of R .

2.2 Factor Models and Spectra

We assume that the observed process w is a q -dimensional, weakly stationary process. A dynamic factor model is a process decomposition of the form $w = \hat{w} + \tilde{w}$, where \tilde{w} is the noise process and \hat{w} is the latent process that is essentially restricted to a linear system. The behaviour \mathcal{B} of \hat{w} is defined as the smallest linear, time invariant, complete system which contains almost all process realizations, that is, $P\{\hat{w}(\omega) \in \mathcal{B}\} = 1$. The following result states that this definition makes sense.

Proposition 1 *For every stochastic process the behaviour is well-defined.*

We call a behaviour nontrivial if $\mathcal{B} \neq (\mathbf{R}^q)^{\mathbb{Z}}$. Dynamic factor models are defined as follows.

Definition 1 *A dynamic factor model of a process w is a decomposition $w = \hat{w} + \tilde{w}$ where the latent process \hat{w} has nontrivial behaviour \mathcal{B} , which is called the behaviour of the factor model.*

In this paper we will be mainly concerned with the behaviour of factor models, as in many cases this is the main point of interest in system identification. In order to simplify our analysis of dynamic factor models we make some additional assumptions on the processes. Some of these assumptions could be relaxed, but they are imposed to prevent technical complications that could obscure the underlying modelling ideas.

Assumptions

- **A1** The processes w , \hat{w} and \tilde{w} are jointly weakly stationary, with zero mean.
- **A2** The observed process w is purely nondeterministic and has full rank.
- **A3** The latent process \hat{w} and the noise process \tilde{w} are purely nondeterministic.
- **A4** The Wold representations of w , \hat{w} and \tilde{w} are absolutely summable.

The assumption A1 is imposed for convenience, as this means that the usual tools of time series analysis and linear systems theory can be applied. The full rank assumption in A2 implies that the behaviour of the observed process is unrestricted, so that it can not be modelled by a factor model without noise. Concerning assumption A3, note that a latent process with nontrivial behaviour can not be of full rank. We assume that it is purely nondeterministic, and that the same holds true for the noise. This seems a reasonable requirement in view of assumption A2. Finally, assumption A4 is imposed for technical reasons. It implies that the spectral densities of the processes are continuous functions on the unit circle.

Stated in terms of behaviours, assumption A3 for the latent process means the following.

Proposition 2 *The behaviour of a purely nondeterministic process is controllable.*

Factor models can also be described by means of spectra. A factor model corresponds to a decomposition

$$\Sigma = \hat{\Sigma} + \tilde{\Sigma} + \Sigma_c + \Sigma_c^* \quad (5)$$

where Σ , $\hat{\Sigma}$ and $\tilde{\Sigma}$ denote the spectrum of w , \hat{w} and \tilde{w} respectively and Σ_c is the cross spectrum between \hat{w} and \tilde{w} . By assumption, the behaviour of the latent process \hat{w} is nontrivial so that $\hat{\Sigma}$ is singular.

2.3 Factor Schemes

In general factor models for a given process w are highly non unique. Here we impose the following alternative restrictions, which we call factor schemes.

- The factor model is called orthogonal if the latent process and the noise process are mutually uncorrelated, that is, if $E\{\hat{w}(t) \tilde{w}(s)'\} = 0$ for all t, s .

- The factor model is called observable if \hat{w} is a linear function of w . Stated otherwise, there holds $\tilde{\Sigma} = F \Sigma F^*$, $\tilde{\Sigma} = (I - F) \Sigma (I - F)^*$ and $\Sigma_c = F \Sigma (I - F)^*$ for some, possibly noncausal, transfer function F .

Even imposing these assumptions does not restrict the class of underlying observationally equivalent systems in the sense that for an arbitrary process w , satisfying assumption A2, every linear system \mathcal{B} may correspond to a factor model. (This may be seen from Theorem 3 below.) So we have to impose additional restrictions on the noise process in order to make meaningful statements about the underlying system. These restrictions should be motivated in each practical situation. The assumption considered here is that the factor model is optimal in the sense explained in the next subsection. For alternative assumptions see e.g. [13].

2.4 Optimal Models

The quality of factor models is expressed in terms of the complexity and the goodness of fit of the model.

Definition 2 *The complexity of a dynamic factor model is defined as the pair (m, n) , where m is the number of driving variables and n the number of states of a minimal state space representation (3) of the behaviour of the factor model.*

The goodness of fit of factor models is measured in terms of the second moments of the noise process \tilde{w} . As is well known, the choice of norms may have an essential effect on the obtained models. Here we will restrict the attention to the mean squares norm and the uniform norm. In the following we use the notation $\tilde{\Sigma}^{1/2}$ for a spectral factor of the noise spectrum $\tilde{\Sigma}$ so that $\tilde{\Sigma} = \tilde{\Sigma}^{1/2} (\tilde{\Sigma}^{1/2})^*$. We define the following norms for spectral factors, where $\lambda_{\max}(Q)$ denotes the spectral radius, that is, the maximum of the absolute values of the eigenvalues of a matrix Q .

$$\|\tilde{\Sigma}^{1/2}\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{trace}\{\tilde{\Sigma}(e^{-i\lambda})\} d\lambda \quad (6)$$

$$\|\tilde{\Sigma}^{1/2}\|_{\infty}^2 = \sup_{\lambda \in [-\pi, \pi]} \lambda_{\max}\{\tilde{\Sigma}(e^{-i\lambda})\} \quad (7)$$

Since the results presented below hold true for both norms we often use the same symbol $\|\tilde{\Sigma}^{1/2}\|$ for both norms. In general, the fit will improve if the model is allowed to be more complex. A factor model is optimal if it optimizes the fit $\|\tilde{\Sigma}^{1/2}\|$ under a given constraint on the complexity (m, n) . The optimality of models depends of course on the specification of the

factor scheme, that is, on the choice of norms for the noise and on possible conditions of orthogonality and observability.

In the following we denote by $\mathbf{B}(m, n)$ the set of all controllable linear systems with m inputs and n states, by $\overline{\mathbf{B}}(m, n) := \bigcup_{k=1}^n \mathbf{B}(m, k)$ the set of all controllable linear systems with m inputs and at most n states, and by $\mathbf{B} := \bigcup_{m=0}^q \bigcup_{n=0}^{\infty} \mathbf{B}(m, n)$ the set of all controllable linear systems.

For given spectrum Σ we denote by $\overline{\mathbf{B}}^*(\Sigma; m, n)$ the set of behaviours of optimally fitting factor models with m inputs and at most n states. Note that $\overline{\mathbf{B}}^*(\Sigma; m, n)$ in general is not a singleton, however we conjecture that in many cases the optimal behaviour will be unique.

First assume that the behaviour of the factor model has been specified a priori, so that the factor equations are given. The aim is to find a model with minimal error that satisfies these equations. Let \mathcal{B} denote the given controllable linear system with polynomial representation $R(\sigma) \hat{w} = 0$ and let $P_{\mathcal{B}}$ be the orthogonal projection defined in (4). The following results hold true both for the mean squares and for the uniform norm.

Theorem 3 *Let w be a process with spectrum Σ and let \mathcal{B} be the prescribed behaviour of a factor model.*

- (i) *A latent process with optimal fit is given by $\hat{w}_0 := P_{\mathcal{B}} w$, with noise spectrum $\tilde{\Sigma}_0 = (I - P_{\mathcal{B}}) \Sigma (I - P_{\mathcal{B}})$. The corresponding factor model is observable, but in general not orthogonal.*
- (ii) *Among orthogonal models, a latent process with optimal fit is given by $\hat{w}_0 := [I - \Sigma R^* (R \Sigma R^*)^{-1} R] w$, with corresponding noise spectrum $\tilde{\Sigma}_0 = \Sigma R^* (R \Sigma R^*)^{-1} R \Sigma$.*

The optimal factor model for a given system \mathcal{B} is unique in case of the mean squares norm but in general not in case of the uniform norm. If we are interested in factor behaviours only, then the above results show that we may restrict the attention to observable models. This leaves four factor schemes of interest, that is, for the mean squares and the uniform norm and according to whether orthogonality is imposed or not. We define the distance between a behaviour and a spectral density as the fit of the optimal factor model with this behaviour. That is, the misfit function is given by

$$d(\Sigma, \mathcal{B}) = \|\tilde{\Sigma}_0^{1/2}\| \quad (8)$$

where $\tilde{\Sigma}_0$ is the noise spectrum of the optimal factor models for \mathcal{B} , given in Theorem 3 and where $\tilde{\Sigma}_0^{1/2}$ denotes a spectral factor of $\tilde{\Sigma}_0$. We use the same notation for the four different factor schemes.

Using the above result the set $\overline{\mathbf{B}}^*(\Sigma, m, n)$ of optimal systems is defined via the minimization of the misfit $d(\Sigma, \mathcal{B})$ over the set $\overline{\mathbf{B}}(m, n)$.

3 Consistency

3.1 System Topology

We introduce the topologies on linear systems and spectra that we will use in our analysis of continuity properties of factor models. For linear systems the gap metric is defined in terms of the projections described in (4).

Definition 3 Let $\mathcal{B}_1, \mathcal{B}_2$ be linear systems, then the gap between these systems is defined by

$$d(\mathcal{B}_1, \mathcal{B}_2) = \|P_{\mathcal{B}_1} - P_{\mathcal{B}_2}\|_\infty \quad (9)$$

In order to investigate continuity properties we also need a topology on the set of spectral densities. We use the metric defined by

$$d(\Sigma_1, \Sigma_2) = \|\Sigma_1 - \Sigma_2\|_\infty \quad (10)$$

Under Assumption A4 the spectra are bounded on the unit circle, so that this is a well-defined metric.

Proposition 4

- (i) The set $\overline{\mathbf{B}}(m, n)$ is the closure of $\mathbf{B}(m, n)$ in \mathbf{B} .
- (ii) The sets \mathbf{B} and $\overline{\mathbf{B}}(m, n)$, for $n > 0$, are not compact.
- (iii) The misfit $d(\Sigma, \mathcal{B})$ is continuous in (Σ, \mathcal{B}) for all positive definite spectra Σ .

3.2 Continuity and Consistency

It is of interest to consider the continuity of optimal systems. Continuity in this respect is connected with robustness, in the sense that small perturbations in the data should lead to a small perturbation of optimal systems.

Proposition 5 The set $\overline{\mathbf{B}}^*(\Sigma, m, n)$ is upper semi continuous in Σ in the sense that $\Sigma_k \rightarrow \Sigma_0$, $\mathcal{B}_k \rightarrow \mathcal{B}_0$ and $\mathcal{B}_k \in \overline{\mathbf{B}}^*(\Sigma_k, m, n)$ implies that $\mathcal{B}_0 \in \overline{\mathbf{B}}^*(\Sigma_0, m, n)$.

Next we investigate the consistency of dynamic factor models when the spectrum is estimated from observed data. In applications the spectrum of the observed process will in general be unknown. Let Σ_T denote an estimator of the process spectrum Σ that is based on a sample $w(1), \dots, w(T)$. Then $\overline{\mathbf{B}}^*(\Sigma_T, m, n)$ is a 'natural' estimate for $\overline{\mathbf{B}}^*(\Sigma, m, n)$. We assume that the estimator is strongly consistent in the sense that that $d(\Sigma, \Sigma_T) \rightarrow 0$ almost surely for $T \rightarrow \infty$.

Theorem 6 The estimator $\overline{\mathbf{B}}^*(\Sigma_T, m, n)$ of the set of optimal systems in $\overline{\mathbf{B}}(m, n)$ is 'upper semiconsistent', in the sense that $\{\mathcal{B}_T \in \overline{\mathbf{B}}^*(\Sigma_T, m, n), \mathcal{B}_T \rightarrow \mathcal{B}_0\} \Rightarrow \{\mathcal{B}_0 \in \overline{\mathbf{B}}^*(\Sigma, m, n)\}$ almost surely.

This means that, under the above conditions, the optimal finite sample systems are in the limit also optimal for the data generating process. However, possibly not all optimal systems are identified in this way.

4 Conclusion

Dynamic factor models decompose an observed process in terms of an underlying latent component and additional noise. The variables are treated in a completely symmetric way, and no assumptions on inputs and outputs are required. The latent process, by assumption, satisfies linear deterministic dynamic relationships. Two different factor schemes are considered. If the noise can be assumed to be uncorrelated with the latent process this is called the orthogonal factor scheme. This is the usual assumption in the classical models of factor analysis. In other situations it is more natural to assume that the factor components are constructed from the observations. This is called the observable factor scheme.

We restrict ourselves to optimal factor models. Here optimal factor models are defined based on the notions of complexity and goodness of fit. Concerning the identification of optimal factor models we present results on consistency.

Several questions deserve further investigation. Of special interest is the further analysis of identification procedures within this framework. Another issue is the incorporation of prior knowledge, for example concerning the input-output structure of the model. A further analysis of the probabilistic structure of factor models is needed in order to develop statistical test procedures, for example to estimate the complexity of factor models from observed data.

References

- [1] T.W. Anderson and H. Rubin, Statistical inference in factor analysis, in J. Neyman (ed.), *Proceedings Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1956, pp. 111-150.
- [2] S. Beghelli, R.P. Guidorzi and U. Soverini, The Frisch scheme in dynamic system identification, *Automatica* 26, 1990, pp. 171-176.
- [3] P.E. Caines, *Linear Stochastic Systems*, Wiley, 1988.

- [4] M. Deistler, Symmetric modeling in system identification, in H. Nijmeijer and J.M. Schumacher (eds.), *Three Decades of Mathematical System Theory*, Springer, 1989, pp. 128-147.
- [5] M. Deistler and W. Scherrer, Identification of linear systems from noisy data, in D. Brillinger et al. (eds.), *New Directions in Time Series Analysis*, part II, IMA vol. 46, Springer, 1992, pp. 21-42.
- [6] M. Deistler and W. Scherrer, System identification and errors in the variables, in K.Haagen et al. (eds.), *Statistical Modelling and Latent Variables*, North-Holland, 1993, pp. 95-111.
- [7] R.F. Engle and M. Watson, A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association* 76, 1981, pp. 774-781.
- [8] R. Frisch, *Statistical Confluence Analysis by means of Complete Regression Systems*, Publ. 5, Economic Institute, University of Oslo, 1934.
- [9] W.A. Fuller, *Measurement Error Models*, Wiley, 1987.
- [10] J.F. Geweke, The dynamic factor analysis of economic time series models, in D.J. Aigner and A.S. Goldberger (eds.), *Latent Variables in Socio-economic Models*, North Holland, 1977, pp. 365-383.
- [11] E.J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, 1988.
- [12] C. Heij, *Deterministic Identification of Dynamical Systems*, Springer, 1989.
- [13] C. Heij, W. Scherrer and M. Deistler, System Identification by Dynamic Factor Models, to appear in *SIAM J. Control Optim.* 35, November 1997.
- [14] R.E. Kalman, A theory for the identification of linear relations, in H. Brezis et al. (eds.), *Colloques Lions*, 1989.
- [15] L. Ljung, *System Identification : Theory for the User*, Prentice-Hall, 1987.
- [16] G. Picci and S. Pinzoni, Dynamic factor analysis models for stationary processes, *IMA Journal of Mathematical Control and Information* 3, 1986, pp. 185-210.
- [17] B. Roorda and C. Heij, Global total least squares modelling of multivariable time series, *IEEE Transactions on Automatic Control* 40, 1995, pp. 50-63.
- [18] J.H. van Schuppen, Stochastic realization problems, in H. Nijmeijer and J.M. Schumacher (eds.), *Three Decades of Mathematical System Theory*, Springer, 1989, pp. 480-523.
- [19] J.C. Willems, From time series to linear system, part I, *Automatica* 22, 1986, pp. 561-580.
- [20] J.C. Willems, From time series to linear system, part III, *Automatica* 23, 1987, pp. 87-115.
- [21] J.C. Willems, Paradigms and puzzles in the theory of dynamical systems, *IEEE Transactions on Automatic Control* 36, 1991, pp. 259-294.