CWI Tract 64

Identifiability, recursive identification and spaces of linear dynamical systems part II

B. Hanzon

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 49E, 49F, 93B, 93D, 93E.
ISBN 90 6196 371 0
NUGI-code: 811

Copyright © 1989, Stichting Mathematisch Centrum, Amsterdam
Printed in the Netherlands

TABLE OF CONTENTS

Part I

	Page
1. INTRODUCTION	6
2. INTRODUCTION TO DETERMINISTIC AND STOCHASTIC LINEAR SYSTEMS	
2.1. Definitions in the deterministic case	12
2.2. Definitions in the stochastic case	22
2.3. Various representations of deterministic finite dimensional linear systems	27
2.3.1. Introduction	27
2.3.2. State space models	27
2.3.3. Arma models	33
2.3.4. The transfer matrix and polynomial mfd (=arma) models	39
2.3.5. Kronecker indices and the specialization order	49
2.4. Stochastic linear systems, various representations	57
2.4.1. Stochastic state space models	57
2.4.2. The spectral density matrix and spectral factorization	63
2.4.3. Hankel matrices of covariances	68
2.4.4. The innovations representation	73
2.4.5. Arma models	73
Appendix 2A. Proof of theorem (2.3.4-3)	75
3. PARTIAL REALIZATION OF ARMA (p,q) MODELS AND THE FINITE IDENTIFIABILITY PROBLEM	
3.1. Partial realization of arma(p,q) models	76
3.1.1. Introduction	76
3.1.2. The deterministic case	77
3.1.2.1. The partial realization lemma	77
3.1.2.2. Unique partial realization of state space models	80
3.1.2.3. The rank structure of the Hankel matrix of arma(p,q) models	82
3.1.3. The stochastic case	95

3.1.3.1. Partial realization of state space models from the covariances 95
3.1.3.2. Partial realization in the stochastic arma $(\bar{p},\bar{q})$ case 96
3.2. The finite identifiability problem 100
3.2.1. On the methodology of the identifiability problem 100
3.2.1.1. A set-theoretic foundation of system identifiability 100
3.2.1.2. Testable stochastic models 106
3.2.1.3. Parameter identifiability 108
3.2.2. Application to the finite identifiability problem 111

4. ON THE DIFFERENTIAL GEOMETRIC AND TOPOLOGICAL STRUCTURE OF FAMILIES OF LINEAR SYSTEMS AND CANONICAL FORMS
4.1. Short introduction 117
4.2. On the definitions of a differentiable manifold and a fibre bundle 117
4.3. Some general remarks on the fibre structure of families of systems-with-states 122
4.4. Nice selections and canonical forms 124
4.5. The differentiable manifold structure of $M^{m}_{m',n,m}$, $M^{m,a}_{m',n,m}$ and $M^{m,a,f}_{m',n,m}$ 128
4.6. The state vector bundle and the associated principal fibre bundle 130
4.7. On the topological structure of some spaces of linear systems 132
4.8. Families of stochastic linear systems 135

5. RIEMANNIAN GEOMETRY AND FAMILIES OF LINEAR SYSTEMS
5.1. Introduction 138
5.2. A Riemannian metric for families of stable deterministic systems 140
5.3. The short time-interval case 156
5.3.1. Introduction 156
5.3.2. A useful family of difference/differential

operators 157
5.3.3. The stability region of the δ-operator and rotations of the Riemann sphere 158
5.3.4. The norm of systems with time interval length Δ 167
5.4. Some results on isometries 174
5.5. Analysis of the Riemannian geometry (inner geometry) in some special, simple cases 185
5.5.1. The Riemannian geometry (inner geometry) of $M_{1,1,1}^{m,a}$ and $M_{1,1,1}^{m,\ell}$ 185
5.5.2. Some other special cases 195
5.6. Degeneration of the Riemannian metric tensor 197
5.7. A Riemannian metric on the state bundle 202
5.8. Riemannian metrics for families of stochastic linear systems 205
5.9. The Fisher metric and Hellinger distance 212
5.10. The Fisher metric on spaces of stochastic systems 215
5.11. Some remarks on identifiability and Riemannian geometry 219
Appendix 5A. Symbolic computation of the Riemannian metric tensors 221

Part II

6. RIEMANNIAN GRADIENT ALGORITHMS FOR RECURSIVE IDENTIFICATION
6.1. Introduction 224
6.2. Description of the algorithm 228
6.2.1. The model set 228
6.2.2. Prediction error algorithms 229
6.2.3. Differences between the new r.p.e. algorithm and the standard one 234
6.2.4. The structure of the coordinate charts and the coordinate changes in the algorithm 235
6.2.5. On the use of local coordinates in the algorithm 239
6.2.6. The Riemannian gradient 241
6.2.7. The transformation rules for a coordinate change 241
6.2.8. Bounding the stepsize 244
6.2.9. The complete set of update equations of the algorithm 247
6.3. A refinement of the cover of the manifold 250
6.3.1. Introduction 250

6.3.2. The structure of a refinement of the n.d.s. cover of the manifold 251
6.3.3. Application to the question of asymptotic stability of products of asymptotically stable matrices 258
6.3.4. Local equivalence of the coordinate chart metrics with the inner metric 261
6.4. On the asymptotic behaviour of the stepsize and the coordinate change times 265
6.5. Spaces of interpolation curves of parameter sequences and their topologies 282
6.6. On a problem of P-a.s. convergence 290
6.6.1. Introduction 290
6.6.2. On (equi-) continuity of some variables in the algorithm 294
6.6.3. On the relation between the coupled and the decoupled algorithm 306
6.7. Exponential decay properties of the algorithm and the implications for convergence 316
6.7.1. About some exponential decay properties 316
6.7.2. Application to the algorithm 329
6.8. The associated differential equation 337
6.8.1. An integral formula for the decoupled algorithm 337
6.8.2. The o.d.e. for the (coupled) algorithm 344
6.9. The asymptotic behaviour of the algorithm 352
6.10. Some final remarks 362
6.10.1. On the question of global convergence to the true parameter point 362
6.10.2. Some remarks on applications and possible extensions of the algorithm 363

Appendix 6A. Computation of $T(\theta)$, $\frac{\partial T(\theta)}{\partial \phi_j^k}$ and $\frac{\partial \phi_i}{\partial \phi_j^T}$. 367

Appendix 6B. Proof of lemma (6.7.2-12) 370

Appendix 6C. On the relationship between V and V_g 373

References 376
List of mathematical symbols 386
Author index 400
Subject index 403
Addenda and errata 411

CHAPTER 6

RIEMANNIAN GRADIENT ALGORITHMS FOR RECURSIVE IDENTIFICATION

6.1. Introduction

In previous chapters we have seen that the set of stochastic linear systems of fixed McMillan degree n, which have no zeroes on the unit circle, forms a differentiable manifold. In chapter 4 we have seen that it is not possible to use only one parametrization.

There are several possibilities to handle this. One approach is to identify the structural indices (Kronecker indices) first and then to identify the parameters within the set of all stochastic systems with those structural indices. (Even then the set of parameters is not a coordinate chart in general, but only some open subset of Euclidean space). Another approach is to use socalled overlapping parametrizations. (cf.[Glo-Wi], [Ove-Lj], [Gui 81], [Cla], i.a.).

Especially for recursive identification this is an important approach because one does not have to decide a priori which structure to fix: one can switch on-line from one parametrization to another. In this chapter we will present an algorithm that uses overlapping parametrizations. It is desirable in such an algorithm that its behaviour does not depend, or at least not very much, on the actual choice of the parametrization at each time. Elsewhere we have started to work out a version of the algorithm that is completely independent of the choice of the parametrization at each time (cf. [Hnz 85b]).

In this chapter we will construct and analyze an algorithm that is (only) asymptotically independent of the choice of parametrization, and apart from that it does not depend very much on the choice of the parametrization at each time, especially if the stepsize is small. We obtain this property by using a socalled Riemannian gradient. As is well-known, on a differentiable manifold a gradient is only defined with respect to a Riemannian metric. It is obtained by premultiplying the gradient in terms of local coordinates (i.e. a chosen parametrization) with the inverse of the Riemannian metric tensor. (see e.g. [Ab-M]). However, for us this is not the basic argument to use a Riemannian gradient. The basic argument is that it has the steepest ascent property: it optimizes the increment of the objective function over all steps of fixed (small) length over all possible directions. The length of such a step should be

measured in the model space, and not in the arbitrary parametrization at hand, because the parametrization is only an instrument to describe the model space, and nothing more. We will try to clarify this argument by way of an example. Consider a (standard) gradient method

$$\hat{\theta}_{t+1} = \hat{\theta}_t - a_t \frac{\partial V}{\partial \theta}, \tag{6.1-1}$$

for a stochastic linear model with parameters θ. The criterion function V depends on the parameters only in so far as V depends on the covariances $\{\Gamma_k\}$ (otherwise the problem would not be identifiable). If a parametrization is such that at certain points, the $\{\Gamma_k\}$ will be (rather) insensitive to certain parameter changes, one can expect in general, that V will also be rather insensitive to such parameter changes. This can be so even if V is not insensitive to changes of the $\{\Gamma_k\}$ in the corresponding directions! As a simple example, consider the following scalar maximum likelihood problem:

$$\begin{cases} x_{t+1} = ax_t + bw_t, \\ y_t = x_t + d.v_t, \end{cases} \tag{6.1-2}$$

v_t, w_t standard white noise. The log-likelihood of the observations $\{y_o, y_1, \ldots, y_T\}$ is equal to

$$V = -\frac{1}{2}\ell n\left|\Gamma(T)\right| - \frac{1}{2}(y_o, y_1, \ldots, y_T)\Gamma(T)^{-1}(y_o, \ldots, y_T)^T, \tag{6.1-3}$$

where

$$\Gamma(T) = \frac{b^2}{1-a^2}\begin{bmatrix} 1 & a & a^2 & \cdots & a^T \\ a & 1 & a & \cdots & a^{T-1} \\ \vdots & & \ddots & & \vdots \\ a^T & \cdots & \cdots & \vdots & 1 \end{bmatrix} + d^2 I_{T+1}. \tag{6.1-4}$$

Then

$$\frac{\partial V}{\partial b} = -\frac{1}{2}tr\left[\Gamma(T)^{-1}\frac{\partial \Gamma(T)}{\partial b}\right] + \\ + \frac{1}{2}(y_o, y_1, \ldots, y_T)\Gamma(T)^{-1}\frac{\partial \Gamma(T)}{\partial b}\Gamma(T)^{-1}(y_o, y_1, \ldots, y_T)^T. \tag{6.1-5}$$

The partial derivative $\frac{\partial V}{\partial d}$ is given by a similar formula. Now consider these formulas at a = 0: then

(6.1-6) $$\frac{\partial \Gamma(T)}{\partial b} = 2bI_{T+1} \text{ and } \frac{\partial \Gamma(T)}{\partial d} = 2dI_{T+1}.$$

Therefore the partial derivative of V in the direction $(\dot{a},\dot{b},\dot{d}) = (0,d,-b)$ at a = 0, is

(6.1-7) $$d\frac{\partial V}{\partial b} - b\frac{\partial V}{\partial d} = 0.$$

It follows that in a neighbourhood of a point $(a_o = 0, b_o, d_o)$, V will be rather insensitive to parameter changes in the direction $(\dot{a},\dot{b},\dot{d}) = (0,d_o,-b_o)$. Therefore the gradient method described above is not likely to go in such a direction even if V increases substantially in such a direction as a function of the covariances. A solution to these problems is obtained by using a Riemannian gradient.

Our basic objective in this chapter is to show that one can construct a recursive identification algorithm on a manifold of stochastic linear systems, which has asymptotic properties similar to those of corresponding algorithms on an open (or at least with non-empty interior) Euclidean parameter space (i.e. $\subseteq \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$). Or to put it concisely our basic objective is to show that 'one can do system identification on a manifold'. As our point of departure we have chosen for Ljung's prediction error algorithm for recursive identification. For a description see e.g. [Lj 81], [Lj-Söd]. We have chosen for the simplest version. Once it is understood how such an algorithm can be transformed to a Riemannian gradient algorithm on a manifold, it should not be very difficult to generalize to less simple versions. To be sure, the simplest version is already quite complicated.
One assumption that we make is rather crucial, especially for the convergence analysis of the algorithm, namely that our manifold is compact (and without boundary in the sense of manifold theory, cf. [Boo], p.11 and around p. 250). For manifolds this assumption is often made, and there are many examples of such manifolds, e.g. a sphere (of any dimension). Although compactness appears to be rather crucial for the algorithm and its convergence properties, this appears not to be the case for the assumption that the manifold has no boundaries. If one has a compact manifold with boundaries, analogous results

are expected. The only problem is that the algorithm for such a case has to be constructed such that if the algorithm reaches the boundary of the manifold, then the next change in parameter is constrained and must not point outside the manifold with boundary. This gives rise to some technical problems that we do not want to go into here. Therefore we assume there are no such boundaries. Because of this, we do not need to have an analogue of the 'projection facility' that is needed in Ljung's prediction error algorithm, and which complicates the algorithm and its analysis (cf. [Lj-Söd]).

The major part of this chapter is taken up by the convergence analysis of the algorithm. One of the reasons for the length of the analysis is that it turned out not to be easily possible (at least for this author) to simply generalize the proofs that exist for the non-manifold case. Instead we had to come up with a new complete proof. It is based on the socalled o.d.e.-method (see e.g. [Lj 77], [Lj-Söd]). In this method of analysis the asymptotic properties of the algorithm are shown to be related to the properties of an ordinary differential equation (o.d.e). To obtain the object that satisfies the o.d.e. we follow the method of [Ku-Cl], (esp. chapter II). Their main idea is to apply a well-known theorem from topological analysis known as the Arzela-Ascoli theorem ,to a set of interpolation curves of the parameterpoints produced by the algorithm. It is a limit point of this set of interpolation curves that satisfies the o.d.e. In [Ku-Cl], p. 19, it is stated that 'the basic idea is simply an extension of the compactness technique as used to construct solutions to ordinary differential equations (cf. [Co-Le], pp. 42-45)'. Kushner and Clark treat some applications of their methods to system identification ([Ku-Cl], pp. 88-98). Once we have established the o.d.e. we can draw rather strong conclusions, thanks to the construction of the solution of the o.d.e. It turns out that our algorithm converges to a compact connected set of critical points of the objective function V_g. Of course the objective function is constant on such a set. This implies that if the critical points of V_g are all isolated, then the algorithm converges to a critical point. In distinction to the theorems of [Lj 77] there are no assumptions needed about the actual behaviour of the algorithm to reach this conclusion. To be more specific: [Lj 77] requires the sequence of parameter points generated by the algorithm to return to a certain set infinitely often, and only under that assumption guaranteed convergence is obtained. We do not need such an assumption.

In section 6.2 the algorithm is presented. In section 6.3 we present a refinement of the cover of the manifold which is needed for the analysis. In section 6.4 the asymptotic behaviour of the stepsizes and the times of coordinate-change is analyzed. In section 6.5 some spaces of interpolation curves are presented, their topological and metrical structure are treated and the relation with the algorithm is explained. One of the main difficulties we found on our way was to deal with probability-one convergence, and the sets of exceptional events of measure zero. All this is discussed in section 6.6. One of the important properties of the algorithm and of the system-to-be-identified is the asymptotic stability of the dynamic matrices involved. Because of this, the behaviour of the algorithm at points of time that lie far apart tend to be almost independent. To make this precise we define certain so-called 'exponential decay' properties, and prove some theorems about them. We hope that these concepts will turn out to be useful for other analyses as well. Together with its implications for convergence of the algorithm, this is treated in section 6.7. In section 6.8 we finally arrive at the associated ordinary differential equation, and in section 6.9 we draw the conclusions that follow from the o.d.e. for the convergence properties of the algorithm. We end the chapter with some final remarks in section 6.10.

6.2. Description of the algorithm

6.2.1. The model set

From theorem (4.8-8) we know that the set of all stochastic systems (4.8-1), with fixed McMillan degree and a fixed number m of output components, which have an innovations representation with asymptotically stable inverse, forms a differentiable manifold, diffeomorphic to $M^{m,a,f}_{m,n,m} \times \text{Pos}(m)$. Fixing the diffeomorphism of theorem (4.8-8), we will identify a stochastic system with the corresponding element of $M^{m,a,f}_{m,n,m} \times \text{Pos}(m)$, if there is no danger of confusion. Our model set will be of the form $M \times \text{Pos}(m)$, with M a compact submanifold of $M^{m,a,f}_{m,n,m}$. To avoid additional technical complications we assume M to be a manifold without boundary (in the sense of manifold theory, cf. e.g. [Boo]). An example of such a space is a sphere (but not a ball). The true model $(\tilde{\theta},\tilde{\Omega})$ will be assumed to lie in the model set:

(6.2.1-1) $\quad (\tilde{\theta},\tilde{\Omega}) \in M \times \text{Pos}(m)$.

Our attention will focus on identifying $\tilde{\theta} \in M$. (In fact, once $\tilde{\theta}$ is identified, one can estimate $\tilde{\Omega}$ by standard procedures, see (6.2.2-4)). Together with M there exists the corresponding (induced) state bundle E and a corresponding principal fibre bundle L. These are defined as follows:

$$(6.2.1\text{-}2) \qquad E := \{\gamma \in M^{m,a,f}_{l,m,n,m} \mid \pi(\gamma) \in M\},$$

$$(6.2.1\text{-}3) \qquad L := \{(A,B,C) \in L^{m,a,f}_{m,n,m} \mid [(A,B,C)] \in M\}$$

(cf. section 4.6, esp. remark 4.6-7).

The manifold M is made into a Riemannian manifold by defining a Riemannian metric on its tangent bundle TM. This can be done as described in chapter 5, although the constructions and results in this chapter hold for an arbitrary Riemannian metric.
The problem to be considered is to construct and analyze a recursive identification algorithm to identify $\tilde{\theta}$ (and $\tilde{\Omega}$) in the model set $M(\times \text{Pos}(m))$.

6.2.2. Prediction error algorithms.
The algorithm that will be constructed is a generalization of the well-known prediction error algorithm (cf. [Lj 81], [Lj-Söd], [Lj 78]). Before describing the generalization in the following subsections, let us briefly review the standard prediction error algorithm.
To be able to apply it, one has to choose, and therefore be able to specify a parameter set $\Theta \subseteq \mathbb{R}^d$, and a smooth mapping

$$(6.2.2\text{-}1) \qquad \Theta \to L^{m,a,f}_{m,n,m}, \quad \theta \longmapsto (A(\theta),B(\theta),C(\theta))$$

with the following two properties:

(i) Θ is an open subset of $\mathbb{R}^d$, or has at least a non-empty interior. If Θ has a non-empty boundary in $\mathbb{R}^d$, the standard prediction error algorithm has built-in a socalled projection facility to see to it that the sequence of parameter estimates that is produced by the algorithm remains within Θ. However, because this will not be needed in our algorithm, we will not go into that here.
(ii) The composition of the mapping (6.2.2-1) with β, i.e. the mapping

$\theta \longmapsto [(A(\theta),B(\theta),C(\theta))] \in M_{m,n,m}^{m,a,f}$, is injective.

If $\theta \in \Theta$ is <u>believed</u> to be the true parameter value, then the <u>corresponding</u> prediction $\hat{y}_t(\theta)$ of y_t will be given by the filter

$$(6.2.2\text{-}2) \quad \begin{cases} x_{t+1}(\theta) = \big(A(\theta)-B(\theta)C(\theta)\big)x_t(\theta) + B(\theta)y_t, \\ \hat{y}_t(\theta) = C(\theta)x_t(\theta). \end{cases}$$

The corresponding <u>prediction error</u> is

$$(6.2.2\text{-}3) \quad \varepsilon_t(\theta) = y_t - \hat{y}_t(\theta).$$

Note that the covariance matrix $\overline{\Omega}$ does <u>not</u> occur in these formulas. If $\theta = \tilde{\theta}$, the true parameter value, then $\overline{\Omega}$ can be estimated consistently by the sample covariance matrix

$$(6.2.2\text{-}4) \quad \hat{\overline{\Omega}} := \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(\tilde{\theta})\varepsilon_t(\tilde{\theta})^T.$$

The idea behind the prediction error algorithm is to try to minimize with respect to θ the expected sum of squares V of the prediction errors:

$$(6.2.2\text{-}5) \quad V(\theta) = \frac{1}{2}E\|\varepsilon_t(\theta)\|^2.$$

From the properties of the steady state Kalman filter it follows that $V(\theta)$ has a unique global minimum at $\theta = \tilde{\theta}$ (cf. [An-M]). So if $V(\theta)$ were known, a method to find the minimum would be the well-known gradient algorithm

$$(6.2.2\text{-}6) \quad \theta_{k+1} = \theta_k - \frac{\partial V}{\partial \theta}(\theta_k),$$

or more generally

$$\theta_{k+1} = \theta_k - R_k(\theta_k)^{-1} \frac{\partial V}{\partial \theta}(\theta_k),$$

where $R_k(\theta_k)$ is some nonsingular (weighting) matrix, usually positive definite symmetric. The gradient is given by

$$(6.2.2\text{-}7)\qquad \frac{\partial V}{\partial \theta} = E\,\frac{\partial \varepsilon_t^T(\theta)}{\partial \theta}.\varepsilon_t(\theta).$$

The expectation, both in (6.2.2-5) and (6.2.2-7), is taken with respect to the true probability measure, which depends on $\tilde{\theta}$ and $\tilde{\Omega}$. They are unknown, so $V(\theta)$ and $\frac{\partial V}{\partial \theta}$ are unknown.
A technique to handle such a situation is the socalled stochastic approximation method (cf. e.g. [Ku-C1], [Lj 81], [Lj-Söd], and the references given there). The idea of this method is to replace (6.2.2-6) by

$$(6.2.2\text{-}8)\qquad \theta_{k+1} = \theta_k - a_k R_k(\theta_k)^{-1}\frac{\partial \varepsilon_t^T(\theta_k)}{\partial \theta}\varepsilon_t(\theta_k),$$

where $\{a_k\}_{k=k_o}^{\infty}$ is a sequence of positive numbers, tending to zero and adding up to infinity. One lets $\{a_k\}$ converge to zero to help 'asymptotically cancel' the noise effects; having the sequence sum to infinity is usually necessary for convergence to the 'right' point or set. ([Ku-C1], p.6). (The role of the parameter t in (6.2.2-8) is perhaps somewhat obscure but we will return to that shortly). From the filter (6.2.2-2), (6.2.2-3), for $\varepsilon_t(\theta)$ one can derive the filter equations for the d × m matrix of partial derivatives $\frac{\partial \varepsilon_t^T(\theta)}{\partial \theta}$. Let θ^i, i = 1,2,...,d, denote the components of the vector θ. One has (compare (6.2.2-2), to keep the notation more transparent we drop the argument θ)

$$(6.2.2\text{-}9)\qquad \begin{cases} x_{t+1} = (A-BC)x_t + By_t, \\ \dfrac{\partial x_{t+1}}{\partial \theta^i} = \dfrac{\partial}{\partial \theta^i}(A-BC).x_t + (A-BC).\dfrac{\partial x_t}{\partial \theta^i} + \dfrac{\partial B}{\partial \theta^i}.y_t, \\ \varepsilon_t = -Cx_t + y_t, \\ \dfrac{\partial \varepsilon_t}{\partial \theta^i} = -\dfrac{\partial C}{\partial \theta^i}.x_t - C.\dfrac{\partial x_t}{\partial \theta^i}. \end{cases}$$

This can be written in vector/matrix notation, as follows. Let

$$\xi_t^T := \left(x_t, \frac{\partial x_t}{\partial \theta^1}, \frac{\partial x_t}{\partial \theta^2}, \ldots, \frac{\partial x_t}{\partial \theta^d}\right),$$ the extended state vector; $\Psi_{it} := \frac{\partial \varepsilon_t}{\partial \theta^i}$, i = 1,2,...,d, then $\left(\varepsilon_t^T, \psi_{1t}^T, \Psi_{2t}^T, \ldots, \Psi_{dt}^T\right)^T$ is the (extended) output vector of

the filter, and $\psi_t^T := [\psi_{1t}\ \psi_{2t}\ \cdots\ \psi_{dt}]^T = \dfrac{\partial \varepsilon_t^T(\theta)}{\partial \theta}$.

Furthermore let

$$(6.2.2\text{-}10)\quad F(\theta) := \begin{bmatrix} A-BC & 0 & \cdots & 0 \\ \dfrac{\partial(A-BC)}{\partial\theta^1} & A-BC & & \vdots \\ \dfrac{\partial(A-BC)}{\partial\theta^2} & 0 & & \\ \vdots & \vdots & & 0 \\ \dfrac{\partial(A-BC)}{\partial\theta^d} & 0 & \cdots\ 0 & A-BC \end{bmatrix},$$

$$(6.2.2\text{-}11)\quad G(\theta) := \begin{bmatrix} B \\ \dfrac{\partial B}{\partial\theta^1} \\ \vdots \\ \dfrac{\partial B}{\partial\theta^d} \end{bmatrix}, \quad K := \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ and}$$

$$(6.2.2\text{-}12)\quad H(\theta) := \begin{bmatrix} -C & 0 & 0 & \cdots & 0 \\ -\dfrac{\partial C}{\partial\theta^1} & -C & 0 & & \vdots \\ -\dfrac{\partial C}{\partial\theta^2} & 0 & -C & \ddots & \\ \vdots & \vdots & & \ddots & 0 \\ -\dfrac{\partial C}{\partial\theta^d} & 0 & \cdots\ 0 & -C & \end{bmatrix}.$$

Note that $F(\theta)$ is asymptotically stable, because $A - BC = A(\theta) - B(\theta)C(\theta)$ is asymptotically stable, for all $\theta \in \Theta$. The filter (6.2.2-9) in vector/matrix notation is:

$$(6.2.2\text{-}13)\quad \begin{cases} \xi_{t+1} = F(\theta)\xi_t + G(\theta)y_t, \\ \begin{bmatrix} \varepsilon_t \\ \psi_{1t} \\ \psi_{2t} \\ \vdots \\ \psi_{dt} \end{bmatrix} = H(\theta)\xi_t + Ky_t. \end{cases}$$

6.2.2-14. Remark. The state space dimension of this filter is n(d+1). In [Gu-Me] it is shown that the McMillan degree of this filter is in general smaller than or equal to n(m+1). So if d > m the representation (6.2.2-13) is not minimal, and could in principle be replaced by a minimal representation. We will not go further into this here.

Now if we take (6.2.2-8) as it stands, then the filter (6.2.2-13) has to be run over and over again, once for each θ_k, to compute $\varepsilon_t(\theta_k)$ and $\psi_t^T(\theta_k)$, for a fixed value of t, and with different input sequences $\{y_\tau\}_{t_o}^{t}$ (preferably independent) in each run. However, we want a recursive procedure, that adapts the parameter estimate on-line. Therefore one proceeds, more or less heuristically, in a 'diagonal' fashion by putting $t \equiv k$ in the formulas. In this way the following algorithm is obtained called the recursive prediction error algorithm

$$
(6.2.2\text{-}15)\quad \begin{cases} \xi_{t+1} = F(\theta_t)\xi_t + G(\theta_t)y_t, \\ \begin{bmatrix} e_t \\ \psi_{1t} \\ \psi_{2t} \\ \vdots \\ \psi_{dt} \end{bmatrix} = H(\theta_t)\xi_t + Ky_t, \\ \hat{\theta}_{t+1} = \hat{\theta}_t - a_t R_t(\theta_t)^{-1}\psi_t e_t, \\ \theta_{t+1} = \hat{\theta}_{t+1}, \end{cases}
$$

with unspecified initial conditions on ξ_{t_o} and $\hat{\theta}_{t_o}$. Note that we have replaced ε_t in (6.2.2-13) by e_t in (6.2.2-15) to make absolutely clear that it is not a prediction error under the hypothesis of some parameter point, but just an auxiliary quantity in the algorithm.

6.2.2-16. Remarks. (i) Usually $\{a_t\}$ is taken such that, besides the conditions mentioned before, it is square summable. The standard example for such a sequence is

(6.2.2-17) $a_t = \frac{\alpha}{(t-t_o)+\beta}$, $\alpha > 0$, $\beta > 0$, $t \geq t_o$.

If the contrary is not explicitly stated, we will assume that $\{a_t\}$ satisfies this extra condition.

(ii) The coupling equation $\theta_t = \hat{\theta}_t$ is introduced here explicitly for two reasons: (a) in the convergence analysis of the algorithm, it will be necessary to 'delete' the coupling equation and to investigate the resulting (data-dependent) map

$$\{\theta_t\}_{t_o}^{\infty} \longmapsto \{\hat{\theta}_t\}_{t_o}^{\infty}.$$

(b) the coupling equation will be generalized (see section 6.2.9).

6.2.3. Differences between the new r.p.e. algorithm and the standard one

In this section we want to list the changes in the algorithm that will be made to obtain what we call a Riemannian gradient recursive prediction error algorithm that uses overlapping parametrizations, or, alternatively, a Riemannian gradient r.p.e. algorithm for manifolds of linear stochastic systems. In the introduction (section 6.2.1) we already mentioned some of the arguments and ideas that are used in the construction of this algorithm.

The equations in the new algorithm deviate from the one presented in the previous section in several respects. Because our parameter space is a manifold, the parameter points will be described by local coordinates. In fact we need local coordinates of the state bundle, because we use the state space in our algorithm. Because of all this the following must be done.

(i) It must be described when and how the algorithm has to change from operating in one coordinate chart to another, and what the topological structure of the coordinate charts has to look like. This will be treated in section 6.2.4.

(ii) The algorithm equations have to be written down in local coordinates. This is treated in section 6.2.5.

(iii) Furthermore, because we work on a manifold, we have to use a more general definition of the concept 'gradient', namely the so-called Riemannian gradient. This is worked out in section 6.2.6.

(iv) If a coordinate change takes place, then a transformation of all relevant variables has to take place. The rules of transformation have to be described. This will be done in section 6.2.7.

(v) For several reasons we need to make sure that the stepsizes of the steps taken in the algorithm are bounded. In practice this will usually be the case, because of the 'physical bounds' of the problem under consideration. Therefore this is mainly a theoretical problem. On the other hand we want to stick to our assumptions and construct the algorithm such that the stepsizes are bounded indeed. We do not want to 'assume away' the problems by making alternative assumptions about the true model. Therefore we are led into an, alas (and hopefully only for the moment) somewhat unelegant way to assure that the stepsizes are bounded. The reason for the way in which this is done is that this affects the probabilistic structure not too much. (Otherwise the proofs would become (even) more complicated). This will be treated in section 6.2.8. (The proof that the procedure presented indeed leads to a bounded stepsize will be given in section 6.4).

Having treated (i) - (v), we will be able to write down the complete set of equations of the new algorithm. This will be done in section 6.2.9.

6.2.4. The structure of the coordinate charts and the coordinate changes in the algorithm.

By definition, any manifold is covered by a set of coordinate charts. In our algorithm we shall make use of such a cover. However, we shall need a cover with a special structure. Making use of the compactness of the manifold M we will be able to show that there will always exist a cover with the required structure.

There are four conditions that we require for the cover $\{C_j | j \in J\}$ of the manifold M. The first condition is that the cover is finite, i.e. $|J| < \infty$. The second one is that it consists of coordinate charts of M, i.e. there are smooth injective coordinate maps $\phi_j: C_j \to \mathbb{R}^d$. The third one is that it consists of coordinate charts of the state bundle over M, or, equivalently, of the corresponding principal bundle L over M, cf. (6.2.1-3). This third condition means that for each chart C_j in the cover, there will exist a smooth, injective coordinate map $\tilde{\phi}_j$ of the bundle L:

$$\tilde{\phi}_j : C_j \times G\ell_n(\mathbb{R}) \to L$$
$$(\theta,T) \longmapsto (TA(\theta,j)T^{-1}, TB(\theta,j), C(\theta,j)T^{-1}).$$

Taking T = I, one obtains a smooth cross section

$$C_j \to L$$
$$\theta \longmapsto (A(\theta,j), B(\theta,j), C(\theta,j)),$$

that will be useful. The fourth condition is somewhat more complicated. We first state a definition.

6.2.4-1. Definition. A set $\{(C_j'', C_j', C_j) \mid j \in J\}$ will be called a nucleus-double-shell (n.d.s.) cover of the manifold M if

(i) C_j'', C_j' and C_j are open for all $j \in J$,

(ii) $\bigcup_{j\in J} C_j'' = M$, $\bigcup_{j\in J} C_j' = M$ and $\bigcup_{j\in J} C_j = M$,

(iii) $\bar{C}_j'' \subseteq C_j' \subseteq \bar{C}_j' \subseteq C_j$ or each $j \in J$.

For each $j \in J$, C_j'' is called the nucleus, $C_j' \setminus C_j''$ is called the first or inner shell and $C_j \setminus C_j'$ is called the second or outer shell.
The fourth condition is that there exists an n.d.s.-cover $\{(C_j'', C_j', C_j) \mid j \in J\}$. Note that because the cover $\{C_j \mid j \in J\}$ satisfies the previous three conditions, the same holds for the covers $\{C_j' \mid j \in J\}$ and $\{C_j'' \mid j \in J\}$.
We will show that there always exists a cover which satisfies these four conditions. We need the following lemma. (This is standard topology and holds in fact for all paracompact manifolds).

6.2.4-2. Lemma. Let $\{C_j \mid j \in J\}$ be an open cover of coordinate charts of the compact manifold M. Then there exists an open cover $\{C_j' \mid j \in J\}$ of M with the property:

(6.2.4-3) $\bar{C}_j' \subseteq C_j \quad \forall j \in J.$

Proof. For each $x \in M$ there exists a $j \in J$ such that $x \in C_j$ and there exists an open neighbourhood $N(x)$ of x, such that $\overline{N(x)} \subseteq C_j$. Of course $\{N(x) | x \in M\}$ is an open cover of M. Because M is compact there exists a finite subcover $\{N(x_k) | k \in K\}$, $|K| < \infty$. Let

$$K_j = \{k | \overline{N(x_k)} \subseteq C_j\},$$

then

$$K = \bigcup_{j \in J} K_j.$$

Let

$$C_j' = \bigcup_{k \in K_j} N(x_k), \text{ then } \bigcup_{j \in J} C_j' = \bigcup_{k \in K} N(x_k) = M.$$

Because each K_j is finite, it follows that

$$\bar{C}_j' = \bigcup_{k \in K_j} \overline{N(x_k)} \subseteq C_j.$$

Q.E.D.

6.2.4-4. Corollary. Let $\{C_j | j \in J\}$ be an open cover of coordinate charts of the compact manifold M. Then there exists an n.d.s. cover $\{(C_j'', C_j', C_j) | j \in J\}$ of M.

Proof. Apply the previous lemma twice.

6.2.4-5. Proposition. Let $M \subseteq M^{m,a}_{m,n,m'}$ be a compact manifold. There exists a cover $\{C_j | j \in J\}$ that satisfies the four condition mentioned before, namely: (i) $|J| < \infty$; (ii) the C_j are coordinate charts of M; (iii) the C_j are coordinate charts of the principal bundle L over M, (iv) there exists an n.d.s. cover $\{(C_j'', C_j', C_j) | j \in J\}$.

Proof. Because M is a manifold there is an open cover of coordinate charts $\{C_\alpha\}$ of M and because L is a (principal) fibre bundle over M, there exists an open cover of bundle-coordinate charts $\{C_\beta\}$ of M. Then the cover $\{C_\alpha \cap C_\beta\}$ of M clearly satisfies (ii) and (iii). Because M is compact there is a finite subcover which we denote by $\{C_j | j \in J\}, |J| < \infty$. This cover satisfies (i), (ii)

and (iii). Apply corollary 6.2.4-4 to obtain a corresponding n.d.s.-cover $\{(C_j'', C_j', C_j) | j \in J\}$. So (iv) is also satisfied.

Q.E.D.

6.2.4-6. Notation. (i) Let $\phi_j: C_j \to \mathbb{R}^d$, $j \in J$, denote coordinate mappings corresponding to the coordinate neighbourhoods. Because $C_j'' \subseteq C_j' \subseteq C_j$, ϕ_j is also defined on C_j'' and C_j'.
(ii) For each point $\theta \in M$, let

(6.2.4-7) $$J''(\theta) := \{j \in J | \theta \in C_j''\}$$

and

(6.2.4-8) $$J'(\theta) := \{j \in J | \theta \in \bar{C}_j'\}.$$

Clearly

$$J''(\theta) \subseteq J'(\theta) \subseteq J.$$

6.2.4-9. Prescription. Let $\theta(r)$, $r \in [a,b)$ be a continuous curve of M. We assign a coordinate chart C_j', with index $j = j(r)$, to each $r \in [a,b)$. We prescribe $j(r)$ to be piecewise constant and left continuous, and $j(a) \in J'(\theta(a))$. A change of coordinates takes place at $r \in [a,b)$ if and only if $\theta(r) \in \partial C_j'$. If so, then $j(r^+) := \lim_{\varepsilon \downarrow 0} j(r+\varepsilon)$, has to be an element of $J''(\theta(r))$.
The prescription is such that a certain 'inner shell' has to be crossed completely between any two coordinate changes. It is clear that this implies that at least a certain fixed positive distance has to be covered between any two coordinate changes. One could call this a form of hysteresis. The procedure can be considered as a generalization of the procedure of [Cla]. For our results it is immaterial which nucleus C_j'' is chosen at a change of coordinates, provided $\theta \in C_j''$ holds. To finish this subsection we give a related proposition for later reference.

6.2.4-10. Proposition. There exists a finite cover $\{U_i\}$ of M with the following property. Let $\theta(r)$, $r \in (a,b) \subseteq \mathbb{R}$, be a continuous curve in M.

Suppose we assign a coordinate chart C_j' with index $j = j(r) \in J(\theta(r))$ to each

r, such that the prescription (6.2.4-9) is satisfied. [I.e. j(r) is piecewise constant and changes at r_o only if $\theta(r_o) \in \partial C_{j(r_o)}$ and changes to

$i \in J'(\theta(r_o))$.] Then, if $(c,d) \subseteq (a,b)$ is such that for some i, $\{\theta(r) | c<r<d\} \subseteq U_i$, i.e. $\theta(r)$ remains within U_i for $r \in (c,d)$, then <u>at most one</u> coordinate change occurs on (c,d).

Proof. Let $\theta \in M$. Consider the compact set

$$\Big(\bigcup_{\theta \notin \partial C'_j} \partial C'_j\Big) \cup \Big(\bigcup_{\theta \notin \partial C''_j} \partial C''_j\Big).$$

Clearly, this set does not contain θ, and therefore there exists an open, (connected) neighbourhood U of θ in the complement of this set. It follows that if, for some $j \in J$, $U \cap C''_j \neq \emptyset$ then (using the connectedness of U) $U \subseteq C''_j \vee \theta \in \partial C''_j$, so $\theta \in \bar{C}''_j$. This implies $\theta \notin \partial C'_j$ and so $U \cap \partial C'_j = \emptyset$. The conclusion is that U has the following property

(6.2.4-11) $\forall j \in J: (U \cap C''_j \neq \emptyset \Rightarrow U \cap \partial C'_j = \emptyset)$.

(Note that θ does not occur in this implication). The sets $U = U(\theta)$, $\theta \in M$ form a cover of M. Because M is compact there exists a finite subcover that will be denoted by $\{U_i | i \in I\}$. Each U_i has the property (6.2.4-11) and therefore if $\exists i \in I: \forall r \in (c,d): \theta(r) \in U_i$, then at most one change of coordinates can occur. Q.E.D.

6.2.5. <u>On the use of local coordinates in the algorithm</u>

First we have to introduce some notation. As described in (6.2.4-6), $\phi_j: C_j \to \mathbb{R}^d$, $j \in J$, denotes the coordinate mapping of the coordinate neighbourhood C_j. If the value of j is clear from the context, we will drop the lower index j and write ϕ. With some abuse of notation, $\phi_j = (\phi_j^1, \phi_j^2, \ldots, \phi_j^d)^T$ will not only denote the mapping, but also the local coordinates themselves. Furthermore, using a similar abuse of notation, we will denote the smooth section defined just before (6.2.4-1), in local coordinates by

$$(6.2.5\text{-}1) \quad \begin{cases} \phi_j(C_j) \ (\subseteq \mathbb{R}^d) \to L, \\ \phi \longmapsto (A(\phi,j), B(\phi,j), C(\phi,j)). \end{cases}$$

Using this, one can take derivatives with respect to $\phi^1,\phi^2,\ldots,\phi^d$ and define $F(\phi,j)$, $G(\phi,j)$, $H(\phi,j)$ (and K) in complete analogy with the definitions (6.2.2-10,11,12). The only changes are that θ has to be replaced by $\phi(\epsilon\mathbb{R}^d)$ and that the index j has to be added to the notation, to replace $F(\theta)$ by $F(\phi,j)$, $A(\theta)$ by $A(\phi,j)$ etc. Corresponding to this, the algorithm-state vector ξ_t is replaced by its local coordinates version $\xi(t,j)$ (and Ψ_t by $\Psi(t,j)$ and $e(t)$ by $e(t,j)$). In fact, $\xi(t,j)$ is the representation in local coordinates of an element of the tangent bundle TE of the state bundle E (in which, as is standard (cf. e.g. [Ko-N]) the state-vector space is identified with its corresponding part of the tangent space TE). This element will be denoted by $\xi_t \epsilon$ TE. This will play a role in the transformation formulas of a coordinate change, that will be treated in section 6.2.7.

Now consider the equations (6.2.2-15). To generalize them to the manifold case, we have to replace the parameter update equation, because the parameter update equation of (6.2.2-15) makes use of addition, which is possible because of the vector space structure of the parameter space there. Two possible solutions to this problem present themselves. (a) One is using the geodesics structure of the manifold (once the Riemannian metric is defined). This is worked out, along with other things, in [Hnz 85b]. The disadvantage of this approach is that it requires propagation of a differential equation in most cases, instead of a simple addition. So it is more complex and it may be computationally burdensome. (b) The other solution is simply to do the addition in the local coordinates. In that case the parameter update equation will be of the form

(6.2.5-2) $\quad \phi_j(\hat{\theta}_{t+1}) = \phi_j(\hat{\theta}_t)$ - "proxy for the gradient",

at least if the right-hand side is an element of $\phi_j(\bar{C}'_j)$. This equation will be worked out further in the next subsections.

6.2.5-3. Remark. In fact, by generalizing the coupling equation $\theta_t = \hat{\theta}_t$, to one of the form

(6.2.5-4) $\quad \theta_t \epsilon B(\hat{\theta}_t,\delta_t)$,

for some sequence of positive numbers $\{\delta_t\}$, (which has to satisfy certain

conditions that will be specified in section 6.2.9), not only the possibility (b) will be captured by the algorithm, but - as it appears - also possibility (a), although this will not be shown here.

6.2.6. The Riemannian gradient

Because the parameter space in our new set-up is a manifold and not a Euclidean space, a more general definition of the concept of the gradient of a function will be used. For the gradient of a real valued differentiable function to be well-defined, one needs a Riemannian metric on the manifold (cf. e.g. [Ab-M], pp. 127-128, especially Def. 2.5.14). In chapter 5 several Riemannian metrics on $M^{m,a}_{m,n,m'}$ were presented. By restriction to $M \subseteq M^{m,a}_{m,n,m'}$, (which is assumed to be a smooth embedding) one obtains a Riemannian metric on M. Which (smooth) Riemannian metric on M is chosen is immaterial for the construction of the algorithm, as presented here, and for the theorems about the convergence behaviour of the algorithm, that will be presented in sections 6.3 - 6.9. With respect to the local coordinates of chart C_j, $j \in J$, the Riemannian metric tensor at a point $\theta \in C_j$ is a positive definite matrix that will be denoted by $R(\theta,j)$. The Riemannian gradient of a differentiable real-valued function V on M is an element of the tangent bundle of M, that is given in local coordinates by

$$(6.2.6\text{-}1) \qquad R(\theta,j)^{-1} \frac{\partial V}{\partial \phi_j},$$

where $\frac{\partial V}{\partial \phi_j} = \left(\frac{\partial V}{\partial \phi_j^1},\ldots,\frac{\partial V}{\partial \phi_j^d}\right)^T$ denotes the vector of partial derivatives of V considered as a function of the local coordinates $(\phi_j^1,\ldots,\phi_j^d)^T \in \phi_j(C_j^o) \subseteq \mathbb{R}^d$.

It is remarkable that in the standard parameter update formula (see (6.2.2-15)), one already finds the expression $R^{-1} \times$ 'proxy for vector of partial derivatives of V'. However, the meaning of the matrix R in the standard case is not completely clear. In the present algorithm R will be taken equal to the Riemannian metric matrix $R(\theta,j)$. In fact, one (and probably more) of the standard choices for R in the literature can be interpreted as a Riemannian metric tensor asymptotically (cf. [Hnz 85a]).

6.2.7. The transformation rules for a coordinate change

A change of coordinates in fact means a change of coordinates of the state

bundle, or equivalently, of the corresponding principal fibre bundle. Therefore such a coordinate change involves the following:

(a) A change in the local coordinates representation of the parameter point $\theta \in M$. In section 6.2.4 it is described which changes of local coordinates are allowed in the algorithm. Suppose the change that takes place is from $\bar{C}'_j$ (so $\theta \in \partial C'_j$) to C''_i (so $\theta \in C''_i$), $j,i \in J$. Then, if $x = \phi_j(\theta) \in \mathbb{R}^d$ denotes the old local coordinate vector, then

$$(6.2.7\text{-}1) \qquad y = \phi_i \circ \phi_j^{-1}(x)$$

is the new one, representing θ in the new local coordinates.
To give an impression of how $\phi_i \circ \phi_j^{-1}$ may look, consider the space $M^{m,a}_{m,n,m'}$. (Of course $M \neq M^{m,a}_{m,n,m'}$, because $M^{m,a}_{m,n,m'}$ is not compact. This is only meant as an impression, nothing more). In sections 4.4 and 4.5 local coordinates are constructed for this space, using a set of local, continuous canonical forms $\{c_\alpha\}_{\alpha \text{ nice}}$. For a coordinate change corresponding to a change from nice selection α to nice selection β the equivalent of the mapping in (6.2.7-1) is the composition of mappings

(6.2.7-2)

$$z \in V^m_\alpha \longmapsto (A_\alpha(z),B_\alpha(z),C_\alpha(z)) \in W^m_\alpha \longmapsto (Q^{-1}A_\alpha(z)Q, Q^{-1}B_\alpha(z), C_\alpha(z)Q) \in W^m_\beta \longmapsto$$

$$(\{R(Q^{-1}A_\alpha(z)Q, Q^{-1}B_\alpha(z))_{s(\beta,j)}\}_{j=1}^m,\ C_\alpha(z)Q) \in V^m_\beta,$$

where $Q = R(A_\alpha(z),B_\alpha(z))_\beta$, invertible, because $z \in V'^m_\alpha \cap V'^m_\beta$ (if not, this coordinate change would of course be impossible). How (6.2.7-1) will look like in detail in our case of the (compact) manifold M will depend on the specific choice of M and on the choice of the coordinate neighbourhoods.
(b) A change of the local section of the principal fibre bundle L of the state bundle $\pi: E \to M$. In other words: instead of the mapping

$$(6.2.7\text{-}3) \quad \begin{cases} C_j \to L, \\ \phi \longmapsto (A(\phi,j),B(\phi,j),C(\phi,j)), \end{cases}$$

the mapping

$$(6.2.7\text{-}4)\quad \begin{cases} C_i \to L, \\ \phi \longmapsto (A(\phi,i),B(\phi,i),C(\phi,i)) \end{cases}$$

is be going to be used.

From (a) we can see that (a) and (b) will be intimately related. In fact, often the local coordinates ϕ_j for the parameter point $\theta \in \phi_j(C_j)$ are certain specified entries of the matrix triple $(A(\theta,j),B(\theta,j),C(\theta,j))$. This is the case for the local coordinates of $M^m_{m,n,m'}$ as specified in chapter 4 (and in (a)). However, this is not necessarily so, and therefore we make this clear distinction.

The effect of the change from (6.2.7-3) to (6.2.7-4) can be described by a nonsingular matrix $T(\theta)$, which describes the state space basis change involved at θ, as follows

(6.2.7-4)

$$(A(\theta,i),B(\theta,i),C(\theta,i)) = (T(\theta)A(\theta,j)T(\theta)^{-1},T(\theta)B(\theta,j),C(\theta,j)T(\theta)^{-1}).$$

To derive the transformation rule for the local coordinates representation of $\xi \in TE$ we proceed as follows. Consider a smooth local section $N(\theta_o) \subseteq M \to E$, $\theta \longmapsto x(\theta)$. Its derivative with respect to θ at θ_o is an element of TE. On the other hand each element of TE can be represented by such a section, in a coordinate independent way. In local coordinates (of the state bundle) $x(\theta)$ can be represented by $x(\phi_j,j) \in \mathbb{R}^n$, $j \in J(\theta)$, (formally together with ϕ_j itself). Its derivative with respect to $\phi_j = (\phi_j^1,\ldots,\phi_j^{d_j})^T$ can be computed and ϕ_j, $x(\phi_j,j)$

and $\dfrac{\partial x(\phi_j,j)}{\partial \phi_j^T}$ together represent an element in TE. The same element of TE is

represented with respect to the coordinates corresponding to C_i

by $\phi_i, x(\phi_i,i)$ and $\dfrac{\partial x(\phi_i,i)}{\partial \phi_i^T}$, where

(6.2.7-5) $\quad x(\phi_i,i) = T(\theta)x(\phi_j,j)$, and so

$$(6.2.7\text{-}6)\quad \frac{\partial x(\phi_i,i)}{\partial \phi_i^k} = \frac{\partial}{\partial \phi_i^k}[T(\theta)].x(\phi_j,j) + T(\theta).\frac{\partial x(\phi_j,j)}{\partial \phi_i^k} =$$

$$= \frac{\partial}{\partial \phi_i^k}[T(\theta)].x(\phi_j,j) + T(\theta).\frac{\partial x(\phi_j,j)}{\partial \phi_j^T}\frac{\partial \phi_j}{\partial \phi_i^k},\ k = 1,2,3,\ldots,d.$$

Because ϕ_i,ϕ_j and $T(\theta)$ are known, the Jacobian $\frac{\partial \phi_j}{\partial \phi_i^T}$ and the derivatives $\frac{\partial T(\theta)}{\partial \phi_i^k}$, $k = 1,\ldots,d$, can be computed. As an alternative to direct computations one can also compute $\frac{\partial T(\theta)}{\partial \phi_i^k}$, $k = 1,\ldots,d$ using Lyapunov equations and the Jacobian. This is treated in appendix 6A. There also a method is given to compute the Jacobian, in case one does not know the transformation mapping $\phi_i \circ \phi_j^{-1}$ explicitly as a function of the parameter.

From (6.2.7-5) and (6.2.7-6) it follows that the vector

$$\left(x(\phi_i,i)^T, \frac{\partial x(\phi_i,i)^T}{\partial \phi_i^1}, \ldots, \frac{\partial x(\phi_i,i)^T}{\partial \phi_i^d}\right)^T \text{ at } \phi_i = \phi_i(\theta), \text{ is a linear}$$

transformation of the vector

$$\left(x(\phi_j,j)^T, \frac{\partial x(\phi_j,j)^T}{\partial \phi_j^1}, \ldots, \frac{\partial x(\phi_j,j)^T}{\partial \phi_j^d}\right)^T$$

Let the matrix of this linear transformation be denoted by $S(\theta;i,j)$. This matrix is completely specified by (6.2.7-5) and (6.2.7-6). Then the transformation rule for $\xi \in TE$ is

$$(6.2.7\text{-}7)\quad \xi(t,i) = S(\theta;i,j)\xi(t,j).$$

6.2.8. Bounding the stepsize

As motivated in section 6.2.3 (v) we will construct the algorithm such that its stepsizes in the parameter space are uniformly bounded. The boundedness of the stepsize will be crucial in our convergence proof for the algorithm. The uniform boundedness is obtained in a perhaps somewhat unelegant way, but it is done such that it does not complicate the probabilistic structure of our algorithm too much.

There are two parts to this socalled boundary provision.

(a) Let $\theta \in M$. For each j such that $\theta \in C_j$, i.e. $j \in J(\theta)$, consider the spectrum of $A(\theta,j)-B(\theta,j)C(\theta,j)$. It is obvious to show that this spectrum is the same for all $j \in J(\theta)$. Indeed this follows directly from the state-space basis change transformation rule (6.2.7-4). Let us denote it by $\sigma(\theta)$, and let $\lambda_M(\theta)$ be an element of $\sigma(\theta)$ with the maximum modulus.

$$(6.2.8\text{-}1) \qquad |\lambda_M(\theta)| = \max_{\lambda \in \sigma(\theta)} |\lambda| < 1.$$

Consider $|\lambda_M(\theta)|$ as a function of $\theta \in M$.

This is a continuous function of θ. This can be shown in three steps.
(i) The coefficients of the characteristic polynomial depend continuously on θ,
(ii) the set $\sigma(\theta)$ of roots of the polynomial depends continuously on its coefficients (cf. [Mar], p.4) and
(iii) $|\lambda_M(\theta)|$ depends continuously on the set $\sigma(\theta)$.
(The details are left to the reader).
Because M is compact, one can define

$$(6.2.8\text{-}2) \qquad \lambda_o := \max_{\theta \in M} |\lambda_M(\theta)| \in (0,1).$$

Choose $\lambda_1 \in (\lambda_o,1)$, and define recursively the nonnegative variables v_t as follows

$$(6.2.8\text{-}3) \qquad \begin{cases} v_{t_o-1} := 0, \\ v_t = \lambda_1 v_{t-1} + \|y_t\|, \ t = t_o, t_o+1, \ldots \end{cases}$$

Note that v_t depends only on the observations y_t and on the choice of λ_1; it does not depend on any quantities computed in the algorithm.
Next choose a ('large') constant $K' > 0$ and define the function

$$(6.2.8\text{-}4) \qquad g_1 : [0,\infty) \to \{0,1\}, g_1(v_t) = 1_{\{v \leq K'\}} = \begin{cases} 1 \text{ if } v_t \leq K', \\ 0 \text{ if } v_t > K'. \end{cases}$$

The idea behind this is to measure with g_1 whether there are outliers in the observations. If $v_t > K'$ then the update equation is 'turned off', and the parameter estimate is kept constant in the algorithm, because otherwise these outliers could destabilize the algorithm. This is comparable to e.g. the method in [Ku-C1], p. 94 (2.6.8), (2.6.9). They have a simpler scheme which we found, however, harder to analyze than ours, because of the dependency on the current parameter estimate.

(b) Let $\{K_t\}_{t_o}^{\infty}$ be a sequence of nonnegative numbers such that

(i) $\lim_{t\to\infty} K_t = \infty$ and

(ii) $\lim_{t\to\infty} a_t K_t = 0$.

($\{a_t\}$ is as in (6.2.2-8) and (6.2.2-16)).

Such a sequence $\{K_t\}_{t_o}^{\infty}$ certainly exists, because $\lim_{t\to\infty} a_t = 0$. For example, one can take $K_t := a_t^{-\frac{1}{2}}$ if $a_t > 0$ and $K_t := t$ if $a_t = 0$.

If the algorithm at time t operates in the j-th coordinate chart, the parameter update equation will be of the following form, if the right-hand side is an element of $\phi_j(\bar{C}_j)$:

$$\text{(6.2.8-5)} \qquad \phi_j(\hat{\theta}_{t+1}) = \phi_j(\hat{\theta}_t) + a_t g_1(v_t) g_2(t) R^{-1} h,$$

with $R = R(\theta_t, j)$ and $h = \Psi(t,j)^T e(t,j)$. The function $g_2(t)$ will be defined as follows:

$$\text{(6.2.8-6)} \qquad g_2(t) = \begin{cases} 1 \text{ if } \|g_1(v_t)R^{-1}h\|_R = g_1(v_t)h^T R^{-1} h \le K_t, \\ 0 \text{ if } \|g_1(v_t)R^{-1}h\|_R = g_1(v_t)h^T R^{-1} h > K_t. \end{cases}$$

Here $\|x\|_R := x^T R x$ denotes the Riemannian length of x, considered as an element of the tangent space TM at θ.

Because $K_t \to \infty$ for $t \to \infty$, the effect of g_2 is vanishing asymptotically. In fact it is only needed to prevent that the algorithm is destabilized if the a_t are too big compared to the sizes of the coordinate charts. Because $\lim_{t\to\infty} a_t = 0$, g_2 plays a role only during a finite time. For more details we

refer to the proof of the uniform boundedness of the stepsize in section 6.4.

6.2.9. The complete set of update equations of the algorithm

Two more things have to be settled before we can write down the complete set of update equations of the algorithm.

(a) If the right-hand side of the parameter update equation (6.2.8-5) lies outside of $\phi_j(\bar{C}_j')$, then $\hat{\theta}_{t+1}$ will be defined to be the point where 'the boundary of $\phi_j(\bar{C}_j')$ is hit' using linear interpolation in the coordinate chart. Define $\lambda_t \in (0,1]$ by

$$(6.2.9\text{-}1) \quad \lambda_t := \min\{1\} \cup \{\lambda \in (0,1) \,|\, \phi_j(\hat{\theta}_t) + \lambda a_t g_1(v_t) g_2(t) R^{-1} h \in \phi_j(\partial \bar{C}_j')\},$$

and let $\hat{\theta}_{t+1}$ be defined by

$$(6.2.9\text{-}2) \quad \phi_j(\hat{\theta}_{t+1}) = \phi_j(\hat{\theta}_t) + \lambda_t a_t g_1(v_t) g_2(t) R^{-1} h.$$

Then $\hat{\theta}_{t+1}$ is well-defined in all cases. Of course if $\lambda_t < 1$ then $\hat{\theta}_{t+1} \in \partial \bar{C}_j'$. If $\hat{\theta}_{t+1} \in \partial \bar{C}_j'$ then a change of coordinates will take place. Note that $\lambda_t \neq 0$ because $\phi_j(\hat{\theta}_t) \in \phi_j(\partial \bar{C}_j')$ in (6.2.9-1). The reason is that if $\phi_j(\hat{\theta}_t) \in \phi_j(\partial \bar{C}_j')$, then a coordinate change will take place immediately, before the parameter update equation is formed.

(b) In (6.2.2-15) the socalled coupling equation $\theta_t = \hat{\theta}_t$ was introduced explicitly. This equation will be generalized as follows. Let $\{\delta_t'\}_{t_o}^{\infty}$ be a sequence of nonnegative numbers that converges to zero and let

$$(6.2.9\text{-}3) \quad \delta_t := a_t \delta_t', \; t = t_o, t_o+1, t_o+2, \ldots .$$

Instead of requiring θ_t to be equal to $\hat{\theta}_t$, we allow θ_t to be chosen arbitrarily from the (nonempty) intersection of

(i) a closed ball $\bar{B}(\hat{\theta}_t, \delta_t)$ with centre $\hat{\theta}_t$ and radius δ_t and

(ii) the set C_j, if $\hat{\theta}_t \in \bar{C}_j'$, and $\bar{C}_j'$ is the coordinate chart in which the algorithm operates at time t. One needs $\theta_t \in C_j$, because one must be able to represent θ_t in the local coordinates $\phi_j = \phi_j(\theta_t)$. So the coupling equation is:

(6.2.9-4) $\theta_t \in \bar{B}(\hat{\theta}_t, \delta_t) \cap C_j$.

Note that, because $\hat{\theta}_t \in \bar{C}_j'$, $\bar{C}_j'$ compact and C_j open, $\bar{C}_j' \subseteq C_j$, it follows that for δ_t small enough

(6.2.9-5) $\bar{B}(\hat{\theta}_t, \delta_t) \subseteq C_j$.

Because δ_t converges to zero, it follows that <u>for the asymptotic analysis</u> one can assume without loss of generality that the coupling equation is

(6.2.9-6) $\theta_t \in \bar{B}(\hat{\theta}_t, \delta_t)$.

For the metric that goes into the definition of $B(\hat{\theta}_t, \delta_t)$ one can choose (i) the inner metric of M, (ii) the Euclidean metric in the local coordinate chart in which the algorithm operates at time t. Those two metrics are, of course locally equivalent around $\hat{\theta}_t$. This will be shown in section 6.3.4. It can be shown that if option (ii) is chosen, there exists a data - independent sequence $\{\tilde{\delta}_t'\}$, (which satisfies the conditions), such that $\{\theta_t, \hat{\theta}_t\}$ satisfies (6.2.9-4) with option (i), too. Therefore without loss of generality in the analysis we will work with option (i), unless otherwise is stated.
Note that one is allowed to take $\delta_t' = 0$ for all $t \geq t_o$. Then the coupling equation reduces to $\theta_t = \hat{\theta}_t$. As an example of how the more general coupling 'equation' can be used, one can take the following parameter update scheme:

(6.2.9-7) $$\theta_{t+1} = \begin{cases} \theta_t & \text{if } \theta_t \in \bar{B}(\hat{\theta}_{t+1}, \delta_{t+1}) \cap C_j^o, \\ \hat{\theta}_{t+1} & \text{if } \theta_t \notin \bar{B}(\hat{\theta}_{t+1}, \delta_{t+1}) \cap C_j^o, \end{cases}$$

where j now indicates the coordinate chart in which the algorithm operates at time t+1. This has the advantage that if the parameter estimate $\hat{\theta}_{t+1}$ is close enough to $\hat{\theta}_t$ then the parameter θ in the computations of the algorithm can be kept <u>constant</u>.
Let us now <u>summarize</u> the update equations of <u>the algorithm</u>

(6.2.9-8) $\xi(t+1,j) = F(\phi,j)\xi(t,j) + G(\phi,j)y_t$;

$$(6.2.9\text{-}9)\qquad \begin{bmatrix} e(t,j) \\ \psi_1(t,j) \\ \vdots \\ \psi_d(t,j) \end{bmatrix} = H(\phi,j)\xi(t,j) + Ky_t,$$

where

$$(6.2.9\text{-}10)\qquad \phi = \phi_j(\theta_t);$$

$$(6.2.9\text{-}11)\qquad \phi_j(\hat{\theta}_{t+1}) = \phi_j(\hat{\theta}_t) + \lambda_t a_t g_1 g_2 R^{-1} h,$$

with λ_t as in (6.2.9-1), g_1 as in (6.2.8-4), g_2 as in (6.2.8-6), and R and h as in (6.2.8-5);

$$(6.2.9\text{-}12)\qquad \theta_{t+1} \in \bar{B}(\hat{\theta}_{t+1},\delta_{t+1}),\ t \geq t_o\text{-}1,\ \text{(cf. (6.2.9-4))}.$$

If $\hat{\theta}_{t+1} \in \partial\bar{C}_j'$ then a coordinate change has to take place to a chart $i \in J$ with the property

$$(6.2.9\text{-}13)\qquad \hat{\theta}_{t+1} \in C_i'',\ \text{i.e. } i \in J''(\hat{\theta}_{t+1});$$

$\phi_j(\hat{\theta}_{t+1})$ has to be replaced by

$$(6.2.9\text{-}14)\qquad \phi_i(\hat{\theta}_{t+1}) = \phi_i \circ \phi_j^{-1}(\phi_j(\hat{\theta}_{t+1})),$$

and $\xi(t+1,j)$ has to be replaced by

$$(6.2.9\text{-}15)\qquad \xi(t+1,i) = S(\theta_{t+1};i,j)\xi(t+1,j).$$

As soon as y_t is known, the calculation (6.2.9-8) - (6.2.9-11) can be made, and the choice (6.2.9-12) can be made. If necessary, a change of coordinates can be made. If all this is done, the algorithm can wait till y_{t+1} becomes available and do all this again, but now with t replaced by t+1 (and if a coordinate change from j to i has taken place, with j replaced by i). This specifies the algorithm except for the initial conditions. Any choice of $j \in J$, $\hat{\theta}_{t_o} \in C_j''$, $\theta_{t_o} \in \bar{B}(\hat{\theta}_{t_o},\delta_{t_o}) \cap C_j$, $\xi(t_o,j) \in \mathbb{R}^{n(1+d)}$ will do in principle. As a standard choice for $\xi(t_o,j)$ one can take $\xi(t_o,j) = 0$, and this

is the case that will be analyzed. If $\xi(t_o,j) \neq 0$, then the definition of v_t has to be changed somewhat. We will not go into this.
This finishes the description of the algorithm, or better, the class of algorithms, because many choices can be made within the class of algorithms: the manifold M, the Riemannian metric, the precise decision rule for a coordinate change (as long as it meets our requirements) and the precise coupling equation (as long as it meets our requirements) etc.

6.3. A refinement of the cover of the manifold

6.3.1. Introduction

'I would rather discover one proof, then to earn the throne of Persia' - Democritos

In the following sections, 6.3-6.10, the asymptotic behaviour of the algorithm presented in section 6.1 and section 6.2 will be analyzed. In the analysis we will make use of a refinement of the n.d.s. cover $\{(C''_j,C'_j,C_j)\}_{j\in J}$ of M. The refinement is introduced only for the sake of the analysis, the algorithm will not be changed. In section 6.3.2 the structure of the refinement, that is needed for our purposes, is described. By choosing a refinement we can make sure that certain properties hold within each coordinate chart. This is applied in section 6.3.3 and section 6.3.4. In section 6.3.3 it is applied to establish asymptotic stability of arbitrary products of dynamic F-matrices occurring in the coordinate neighbourhood involved. In section 6.3.4 it is applied to establish equivalence in each coordinate neighbourhood between the inner metric of the Riemannian manifold and the metric defined by using the Euclidean metric of the local coordinates.

6.3.2. The structure of a refinement of the n.d.s. cover of the manifold

The refinements we will consider will again have a "nucleus-double-shell" (n.d.s.) structure just as the cover $\{(C_j'', C_j', C_j) | j \in J\}$. Consider a finite open cover $\{E_i | i \in I\}$ of M, $|I| < \infty$.

Let

$$E_{ij} := E_i \cap C_j \text{ for all } i \in I,\ j \in J. \tag{6.3.2-1}$$

Then $\{E_{ij}\}$ forms a finite open cover of M and for fixed j, $\{E_{ij} | i \in I\}$ forms a finite open cover of C_j, and therefore $\bar{C}_j' \subseteq \bigcup_{i \in I} E_{ij}$.

6.3.2-2. Proposition. Let $\{E_{ij} | i \in I, j \in J\}$ be any finite open cover of M with the property $\bar{C}_j' \subseteq \bigcup_{i \in I} E_{ij}$. Then there exists an n.d.s. - cover

$\{(E_{ij}'', E_{ij}', E_{ij}) | i \in I,\ j \in J\}$ of M such that

$$\bigcup_{i \in I} E_{ij}'' = C_j', \quad \forall j \in J. \tag{6.3.2-3}$$

6.3.2-3. Remark. This is also standard topology; note that some of the E_{ij}'', E_{ij}' and E_{ij} may be empty sets. Let

$$I(j) := \{i \in I | E_{ij}'' \neq \emptyset\}. \tag{6.3.2-4}$$

Then $\{(E_{ij}'', E_{ij}', E_{ij}) | i \in I(j), j \in J\}$ will again be an n.d.s. cover of M!

Proof of proposition (6.3.2-2) (sketch).

This is analogous to the proofs of lemma 6.2.4-2 and corollary 6.2.4-4. First one shows that if $\{E_{ij} | i \in I\}$ covers $\bar{C}_j'$ in the sense that

$$\bar{C}_j' \subseteq \bigcup_{i \in I} E_{ij} \tag{6.3.2-5}$$

then there exists a nucleus-shell (n.s.) cover $\{(E_{ij}', E_{ij}) | i \in I\}$ of $\bar{C}_j'$, analogous to lemma (6.2.4-2) and its proof, using the fact that $\bar{C}_j'$ is compact. This means that $\bar{E}_{ij}' \subseteq E_{ij}, \forall i \in I$ and that

$$\bar{C}_j' \subseteq \bigcup_{i \in I} E_{ij}'. \tag{6.3.2-6}$$

Then by applying the same argument to $\{E'_{ij}\}$, (compare corollary (6.2.4-4)) one concludes that for each $j \in J$ there exists also an open n.d.s. cover $\{(\tilde{E}''_{ij}, E'_{ij}, E_{ij}) | i \in I\}$ such that

(6.3.2-7) $\quad \bar{C}'_j \subseteq \bigcup_{i \in I} \tilde{E}''_{ij}.$

Now let $E''_{ij} := \tilde{E}''_{ij} \cap C'_j$. Then $\bar{E}''_{ij} \subseteq E'_{ij}$ clearly. It follows furthermore that

(6.3.2-8) $\quad \bigcup_{i \in I} E''_{ij} = C'_j$

and

(6.3.2-9) $\quad \bigcup_{j \in J} \bigcup_{i \in I} E''_{ij} = M$

and so $\{(E''_{ij}, E'_{ij}, E_{ij}) | i \in I, j \in J\}$ is indeed an open n.d.s. cover of M with the required property. Q.E.D.

Now let for all $i \in I(j)$, all $j \in J$

(6.3.2-10) $\quad D''_{ij} = E''_{ij}, \; D'_{ij} = E'_{ij} \cap C'_j, \; D_{ij} = E_{ij} \cap C'_j.$

6.3.2-11. Proposition. (a) $\{(D''_{ij}, D'_{ij}, D_{ij}) | i \in I(j), j \in J\}$ has the following properties

(i) $\quad D''_{ij}, D'_{ij}, D_{ij}$ are open sets for each i,j,

(ii) $\quad D''_{ij} \subseteq D'_{ij} \subseteq D_{ij}$ holds for all i,j,

(iii) $\quad \bigcup_{i \in I(j)} D''_{ij} = C'_j$ for all $j \in J$ and

(iv) for each $i \in I(j)$, $j \in J$ there exist open sets N'_{ij}, $N_{ij} \subseteq C_j$, such that

(6.3.2-12) $\quad \bar{D}''_{ij} \subseteq N'_{ij}$ and $N'_{ij} \cap C'_j \subseteq D'_{ij}$, and
$\quad \bar{D}'_{ij} \subseteq N_{ij}$ and $N_{ij} \cap C'_j \subseteq D_{ij}$.

(b) If $\{(D''_{ij},D'_{ij},D_{ij})\,|\,i \in I(j),\ j \in J\}$ has the properties (i), (ii), (iii), (iv) mentioned in (a), then there exists an n.d.s. cover $\{(E''_{ij},E'_{ij},E_{ij})\,|\,i \in I(j), j \in J\}$ of M, such that for all $i \in I(j)$, $j \in J$:

(1) $E_{ij} \subseteq C_j$,

(2) $E''_{ij} = D''_{ij}$; $E'_{ij} = D'_{ij} \cap C'_j$; $E_{ij} = D_{ij} \cap C'_j$.

Before we go to the proof of this proposition we give a definition and some remarks.

6.3.2-13. Definition. $\{(D''_{ij},D'_{ij},D_{ij})\,|\,i \in I(j), j \in J\}$ having properties (i), (ii), (iii), (iv) of the previous proposition, will be called a second order n.d.s. cover.

6.3.2-14. Remarks (i) It is called a second order n.d.s. cover because it is a refinement of the n.d.s. cover $\{(C''_j,C'_j,C_j)\,|\,j \in J\}$ with respect to the n.d.s. cover $\{(E''_{ij},E'_{ij},E_{ij})\,|\,i \in I(j),\ j \in J\}$.
(ii) $\{(D''_{ij},D_{ij},D_{ij})\,|\,i \in I(j), j \in J\}$ is itself not an n.d.s. cover because

$\bar{D}''_{ij} \subseteq D'_{ij}$ etc. does not necessarily hold. On the other hand for all $i \in I(j)$, $j \in J$:

(6.3.2-15) $\bar{D}''_{ij} \cap C'_j \subseteq D'_{ij}$, $\bar{D}'_{ij} \cap C'_j \subseteq D_{ij}$.

However, these inclusion are not sufficient for a second order n.d.s. cover. The reason behind it is that (6.3.2-15) does not exclude shells with 'thickness' vanishing at certain points, while definition (6.3.2-13) does exclude this, as will be shown after the proof of the proposition.

Proof of proposition 6.3.2-11.

(a) (i), (ii), (iii) are trivial; (iv) can be shown simply by taking $N'_{ij} := E'_{ij}$ and $N_{ij} := E_{ij}$, for all $i \in I(j)$, all $j \in J$.

(b) Let $C_j^{\frac{1}{2}}$ be an open set such that $\bar{C}'_j \subseteq C_j^{\frac{1}{2}} \subseteq \bar{C}_j^{\frac{1}{2}} \subseteq C_j$. Let N'_{ij} and N_{ij} be as in (6.3.2-12). Because $\bar{D}''_{ij} \subseteq N'_{ij}$, there exists an open set $N_{ij}^{\frac{1}{2}}$ such that

$\bar{D}''_{ij} \subseteq N^{\frac{1}{2}}_{ij} \subseteq \bar{N}^{\frac{1}{2}}_{ij} \subseteq N'_{ij}$. Define (for all $i \in I(j)$, $j \in J$)

$$E''_{ij} := D''_{ij},$$

$$E'_{ij} := N^{\frac{1}{2}}_{ij} \cup D'_{ij} \text{ and}$$

$$E_{ij} := N'_{ij} \cup N_{ij} \cup D_{ij}.$$

Then

$$\bar{E}''_{ij} = \bar{D}_{ij} \subseteq N^{\frac{1}{2}}_{ij} \subseteq E'_{ij} \text{ and}$$

$$\bar{E}'_{ij} = \bar{N}^{\frac{1}{2}}_{ij} \cup \bar{D}'_{ij} \subseteq N'_{ij} \cup N_{ij} \subseteq E_{ij}.$$

Furthermore, one has:

(6.3.2-16) $E''_{ij} \cap C'_j = D''_{ij}$,

because $E''_{ij} = D''_{ij} \subseteq C'_j$;

(6.3.2-17) $E'_{ij} \cap C'_j = D'_{ij}$,

because $E'_{ij} = N^{\frac{1}{2}}_{ij} \cup D'_{ij}$ and $N^{\frac{1}{2}}_{ij} \cap C'_j \subseteq N'_{ij} \cap C'_j \subseteq D'_{ij} \subseteq C'_j$;

(6.3.2-18) $E_{ij} \cap C'_j = D_{ij}$,

because $E_{ij} = N'_{ij} \cup N_{ij} \cup D_{ij}$ and $N'_{ij} \cap C'_j \subseteq D'_{ij} \subseteq N_{ij}$ and

$N_{ij} \cap C'_j \subseteq D_{ij} \subseteq C'_j$.

Q.E.D.

For a cover with this structure we have the following important property.

6.3.2-19. Proposition. There exists a constant $c > 0$ such that for all $j \in J$ and for all $i \in I(j)$, one has

(6.3.2-20) $d(D''_{ij}, \partial D'_{ij} \setminus \partial C'_j) \geq c$,

where the distance d between sets A and B is defined as

$$d(A,B) = \inf_{a\in A, b\in B} d(a,b) \text{ and } d(A,B) = \infty \text{ if } A = \emptyset \text{ or } B = \emptyset.$$

Proof. Because enlarging the sets can not increase the distance, one has

$$d(D''_{ij}, \partial D'_{ij} \setminus \partial C'_j) \geq d(\bar{E}''_{ij}, \partial E'_{ij}),$$

(the $\{(E''_{ij}, E'_{ij}, E_{ij})\}$ are as in proposition 6.3.2-11(b)). Here it is used that

$D''_{ij} = E''_{ij}$ and $D'_{ij} = C'_j \cap E'_{ij}$ which implies $\partial D'_{ij} \subseteq \partial C'_j \cup \partial E'_{ij}$, and so

$$\partial D'_{ij} \setminus \partial C'_j \subseteq \partial E'_{ij} \setminus \partial C'_j \subseteq \partial E'_{ij}.$$

Because $\bar{E}''_{ij} \subseteq E'_{ij}, E'_{ij}$ open, it follows that $\bar{E}''_{ij} \cap \partial E'_{ij} = \emptyset$. Furthermore $\bar{E}''_{ij}$ and $\partial E'_{ij}$ are closed and therefore compact, so $d(\bar{E}''_{ij}, \partial E'_{ij})$ is positive, say:

$d(\bar{E}''_{ij}, \partial E'_{ij}) = d_{ij} > 0$. Now let $c = \min_{\substack{i\in I(j) \\ j\in J}} d_{ij} > 0$.

Q.E.D.

For the sake of the analysis of the algorithm we will associate not only a coordinate chart C_j, but also a subchart $D_{ij} \subseteq C_j$, with each stage of the algorithm. The rules of changing from D_{ij} to D_{kl} are a generalization of those of changing from C_j to C_ℓ. Furthermore, <u>for the sake of the analysis</u>, we will generalize the prescription somewhat, such that <u>curves</u> that enter $\partial \bar{D}'_{ij}$ but do not <u>leave</u> $\bar{D}'_{ij}$ are <u>allowed</u> to go <u>without</u> a change of coordinates. One could generalize the prescription (6.2.4-9) likewise, in which case the algorithm itself would be generalized. However, for the sake of definiteness of the rule for changing coordinate charts in the algorithm, we choose not to do so.

6.3.2-21. <u>Prescription</u>. (Compare (6.2.4-9))
Let $\theta(t)$, $r \in [a,b)$ be a continuous curve of M. We assign a coordinate chart $\bar{D}'_{ij}$ with index pair $(i,j) = (i(r),j(r))$, to each $r \in [a,b)$. We prescribe $(i(r),j(r))$ to be piecewise constant and left continuous and $(i(r),j(r))$ has to be such that

$$\forall r \in [a,b): \theta(r) \in \bar{D}'_{i(r),j(r)}.$$

A change of coordinates is allowed to take place at $r \in [a,b)$ only if

$$\theta(r) \in \partial\bar{D}'_{i(r),j(r)}.$$

If it takes place then $(i(r^+),j(r^+)) := \lim_{\varepsilon\downarrow o}(i(r+\varepsilon),j(r+\varepsilon))$ has to be such that

$$\begin{cases} \text{if } \theta(r) \in \partial''C'_{j(r)} \text{ then } j(r^+) \text{ such that } \theta(r) \in C''_{j(r^+)} \text{ and } i(r^+) \text{ such that} \\ \quad \theta(r) \in D''_{i(r^+),j(r^+)}, \\ \text{if } \theta(r) \in \partial''D'_{i(r),j(r)} \setminus \partial''C'_{j(r)}, \text{ then } j(r^+) = j(r) \text{ and } i(r^+) \text{ such that} \\ \quad \theta(r) \in D''_{i(r^+),j(r^+)}. \end{cases}$$

It is very important that for $j(r)$ this prescription is a generalization of the one given before in (6.2.4-9). (Note that $\partial''C'_j \subseteq \bigcup_{i\in I(j)} \partial''D'_{ij}$). To conclude this subsection we state the analog of (6.2.4-10).

6.3.2-22. Proposition. There exists a finite cover $\{U_k\}$ of M with the following property. Let $\theta(t)$, $t \in (a,b)$ be a continuous curve in M. If an interval $(c,d) \subseteq (a,b)$ is such that for some value of k and for all $t \in (c,d)$, $\theta(t) \in U_k$, then at most two coordinate changes occur on the interval (c,d). I.e. $(i(r),j(r))$ takes on at most three values for $r \in (c,d)$.

Proof. Let $\{(E''_{ij},E'_{ij},E_{ij}) \mid i \in I(j), j \in J\}$ be an n.d.s. cover of M such that $D''_{ij} = E''_{ij} \cap C'_j$; $D'_{ij} = E'_{ij} \cap C'_j$ and $D_{ij} = E_{ij}$, as before. Let $\theta \in M$. Consider the compact set

$$\Big(\bigcup_{\theta\notin\partial''C'_j} \partial''C'_j\Big) \cup \Big(\bigcup_{\theta\notin\partial C''_j} \partial C''_j\Big) \cup \Big(\bigcup_{\theta\notin\partial''E'_{ij}} \partial''E'_{ij}\Big) \cup \Big(\bigcup_{\theta\notin\partial E''_{ij}} \partial E''_{ij}\Big).$$

Clearly this set does not contain θ, and therefore there exists an open connected neighbourhood $U = U(\theta)$ of θ in the complement of this set. Just as in the proof of (6.2.4-10) U has property (6.2.4-11):

(6.3.2-23) $\forall j \in J$: $U \cap C_j'' \neq \emptyset \Rightarrow U \cap \partial C_j' = \emptyset$

and similarly

(6.3.2-24) $\forall i \in I(j)$, $j \in J$: $U \cap E_{ij}'' \neq \emptyset \Rightarrow U \cap \partial E_{ij}' = \emptyset$.

We claim that if $\theta(r)$, $r \in [a,b)$ is a continuous curve which takes its values in U, then at most two coordinate changes (in the sense of the prescription (6.3.2-21)) can take place. According to the prescription, after a first coordinate change at $r = r_1$ (say) we have $\theta(r_1^+) \in D_{ij}''$ with $i = i(r_1^+)$, $j = j(r_1^+)$. Because $\theta(r_1^+) \in U$ and $D_{ij}'' = E_{ij}''$, (6.3.2-24) implies that $U \cap \partial E_{ij}' = \emptyset$. A second coordinate change can only take place at $\partial D_{ij}' \cap U$. We know that $\partial D_{ij}' \subseteq \partial C_j' \cap \partial E_{ij}'$. It follows that

$$\partial D_{ij}' \cap U \subseteq \partial C_j' \cap U \subseteq \partial C_j'.$$

Therefore a second coordinate change at r_2 can only take place if $\theta(r_2) \in \partial C_j'$. In that case j will be changed. According to the prescription we will have

$$\theta(r_2) \in D_{k\ell}'' \cap C_\ell'' \text{ with } k = i(r_2^+),\ \ell = j(r_2^+).$$

It then follows from (6.3.2-23), (6.3.2-24) that

$$U \cap \partial C_j' = \emptyset \text{ and } U \cap \partial E_{ij}' = \emptyset.$$

So $U \cap \partial D_{ij}' \subseteq U \cap (\partial C_j' \cup \partial E_{ij}') = \emptyset$.
Therefore after a second coordinate change no more coordinate changes will take place for $\theta(r)$.
It is clear that $\{U(\theta) \mid \theta \in M\}$ is an open cover of M. Because M is compact there is a finite subcover $\{U_k\}$ of M, and each U_k has the required property.

Q.E.D.

6.3.2-25. Remark. According to the proof $\{U_k\}$ can be chosen such that if two coordinate changes take place then the first one will leave j constant, while the second one changes j. I.e. the first one will be within C_j', while the second one will be a coordinate change from $\bar{C}_j'$ to C_ℓ''.

6.3.3. Application to the question of asymptotic stability of products of asymptotically stable matrices

Let us consider the problem of possible instability of a product of asymptotically stable matrices. If A_1 and A_2 are asymptotically stable matrices, then the product does not have to be asymptotically stable. For example, let

$$(6.3.3\text{-}1) \quad A_1 = \begin{pmatrix} 0.5 & 0 \\ 10 & 0.5 \end{pmatrix}, \; A_2 = A_1^T.$$

Then

$$(6.3.3\text{-}2) \quad A_1A_2 = \begin{pmatrix} 0.25 & 5 \\ 5 & 100.25 \end{pmatrix}$$

which is unstable.

This implies that a time varying linear system

$$(6.3.3\text{-}3) \quad \begin{cases} x_{t+1} = A_t x_t + B_t u_t \\ y_t = C_t x_t \end{cases}$$

with A_t an asymptotically stable matrix for each t (i.e. $\forall t$: $\sigma(A_t) \subseteq D(0,1)$), can be unstable. An example is obtained by taking $A_{2t+1} := A_1$ and $A_{2t} := A_2$, for all t, with A_1 and A_2 as above.

In our algorithm the dynamic matrix is F_t, which is asymptotically stable (for each t) but time-varying. So the question arises whether the resulting time-varying system is stable. To treat this problem we will investigate the relationship between asymptotic stability and asymptotically stable norms.

6.3.3-4. Definition (cf. [Gan]).

The right norm of a matrix A is AA^* and the left norm A^*A.

It is well-known that AA^* and A^*A have the same nonzero eigenvalues (they are the squares of the nonzero singular values of A); and we will say that A has asymptotically stable norm(s) if AA^* and A^*A are asymptotically stable, i.e. if their eigenvalues are smaller than one, or, equivalently, if all the singular values of A are smaller than one. Clearly a has asymptotically stable norm if and only if $\|A\|_S < 1$ ($\|,\|_S$ denotes again the spectral norm).

6.3.3-5. Proposition. Let A be square and let λ_M denote the eigenvalue of A with maximum modulus:

$$|\lambda_M| = \max_{\lambda_i \in \sigma(A)} |\lambda_i|.$$

Then

(a) $|\lambda_M| \leq \|A\|_S$.

(b) For each $\varepsilon > 0$ there exists a nonsingular matrix T and an open neighbourhood N of A such that for all $\tilde{A} \in N$

(6.3.3-6) $$\|T\tilde{A}T^{-1}\|_S < |\lambda_M| + \varepsilon.$$

Before going to the proof let us state a corollary.

6.3.3-7. Corollary. Let A be square.
(a) If A has asymptotically stable norm then A is asymptotically stable.
(b) If A is asymptotically stable then there exists a nonsingular matrix T and an open neighbourhood N of A such that for all $\tilde{A} \in N$, $T\tilde{A}T^{-1}$ has asymptotically stable norm. (and especially TAT^{-1} has asymptotically stable norm).

Proof of proposition (6.3.3-5). (a) Let $Ax_0 = \lambda_M x_0$, $\|x_0\| = 1$, $x_0 \in \mathbb{C}^n$ then it follows that $\|A\|_S = \max_{\|x\|=1} \|Ax\| \geq \|\lambda_M x_0\| = |\lambda_M|$.

(b) Because $\|T\tilde{A}T^{-1}\|_s$ is continuous in the entries of $\tilde{A}$, it suffices to show that for each $\varepsilon > 0$ there exists a nonsingular matrix T such that
$\|TAT^{-1}\|_S < |\lambda_M| + \varepsilon$.
Let S be an invertible matrix such that

(6.3.3-8) $$J_1 := SAS^{-1}$$

is in Jordan normal form. Then J_1 is block-diagonal with $n_i \times n_i$ blocks $J_1^{(i)}$, $i = 1,2,\ldots,I_o$ (say), of the form

(6.3.3-9) $$J_1^{(i)} = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}.$$

Let J_μ, with $\mu > 0$, be defined by

(6.3.3-10) $J_\mu := \Lambda_\mu^{-1} J_1 \Lambda_\mu$,

where $\Lambda_\mu := \mathrm{diag}(1,\mu,\mu^2,\ldots,\mu^{n-1})$. Then J_μ has the same block structure as J_1, with as its i^{th} block

$$J_\mu^{(i)} = \begin{bmatrix} \lambda_i & \mu & & 0 \\ & \lambda_i & \ddots & \\ & & \ddots & \mu \\ 0 & & & \lambda_i \end{bmatrix}.$$

Clearly $\lim_{\mu \downarrow o} J_\mu^{(i)} = \mathrm{diag}(\lambda_i,\lambda_i,\ldots,\lambda_i)$, and $J_o := \lim_{\mu \downarrow o} J_\mu$ is a diagonal matrix with the eigenvalues of A on its main diagonal. Therefore $\|J_o\|_S = |\lambda_M|$. Because the spectral norm depends continuously on the entries of the matrix, it follows that for each $\varepsilon > 0$ there exists a number $\mu_o > 0$ such that $\|J_{\mu_o}\|_S < |\lambda_M| + \varepsilon$. So let $T = \Lambda_{\mu_o}^{-1} S$ and (b) follows.

Q.E.D.

This will now be applied, as follows. Let $\theta \in M$ and j such that $\theta \in C_j$. Consider $F(\phi_j(\theta),j)$ as defined in section 6.2. Its eigenvalues are the same as those of $A(\theta,j) - B(\theta,j)C(\theta,j)$ (compare (6.2.2-10)), but with higher multiplicity. Therefore its eigenvalue with maximum modulus is $\lambda_M(\theta)$ (cf. (6.2.8-1)), and $\forall\theta \in M$: $|\lambda_M(\theta)| \leq \lambda_o < 1$ (cf. (6.2.8-2)). Let $\varepsilon \in (0,1-\lambda_o)$ be fixed. According to the previous proposition there exists a neighbourhood $N \subseteq C_j$ of θ and a (constant, nonsingular) matrix T such that

$$\forall\tilde{\theta} \in N: \|TF(\phi_j(\tilde{\theta}),j)T^{-1}\|_S < \lambda_o + \varepsilon \ (< 1).$$

The neighbourhoods cover M. Because each $\bar{C}_j'$ is compact and $|J| < \infty$, there is a finite subcover $\{E_{\alpha j}\}$ with the property $\bar{C}_j' \subseteq \bigcup_\alpha E_{\alpha j}$. At the end of the next subsection this cover will be used to obtain a second order n.d.s. cover $\{(D_{ij}'',D_{ij}',D_{ij}) \mid i \in I(j),\ j \in J\}$ corresponding to $\{(C_j'',C_j',C_j) \mid j \in J\}$, such that for each set D_{ij} there exists a nonsingular matrix T_{ij} such that for all $\theta \in D_{ij}$

(6.3.3-11) $F(\phi_j(\theta);i,j) := T_{ij}F(\phi_j(\theta),j)T_{ij}^{-1}$

has spectral norm smaller than $\lambda_o + \varepsilon$. In accordance with this we will define (compare (6.2.9-8) and (6.2.9-9))

(6.3.3-12) $G(\phi_j(\theta);i,j) = T_{ij}G(\phi_j(\theta),j)$

and

(6.3.3-13) $H(\phi_j(\theta);i,j) = H(\phi_j(\theta);i,j)T_{ij}^{-1}$

and

(6.3.3-14) $\xi(t;i,j) = T_{ij}\xi(t,j)$.

Equations (6.2.9-8) and (6.2.9-9) can then be rewritten as

(6.3.3-15) $\xi(t+1;i,j) = F(\phi;i,j)\xi(t;i,j) + G(\phi;i,j)y_t$,

(6.3.3-16) $$\begin{bmatrix} e(t,j) \\ \Psi_1(t,j) \\ \vdots \\ \Psi_d(t,j) \end{bmatrix} = H(\phi;i,j)\xi(t;i,j) + Ky_t,$$

if $\phi = \phi_j(\theta)$, where $\theta \in D_{ij}$.

6.3.4. Local equivalence of the coordinate chart metrics with the inner metric

(This is of course standard differentiable geometry, included for completeness sake). Within a coordinate chart (C_j,ϕ_j) one can make use of the Euclidean metric d_j of the coordinates. To be precise, d_j is given by:

(6.3.4-1) $d_j(\theta_o,\theta_1) := \|\phi_j(\theta_o)-\phi_j(\theta_1)\|$ for all $\theta_o,\theta_1 \in C_j$.

The length of a differentiable curve γ: $[0,1] \to C_j$, in this metric, is given by the formula

$$(6.3.4\text{-}2) \qquad I_j(\gamma) = \int_0^1 \left\| \frac{d\phi_j(\gamma(t))}{dt} \right\| dt.$$

Let $\Gamma(\theta_o,\theta_1,U)$ denote the set of all differentiable curves γ: $[0,1] \to U$ with $\gamma(0) = \theta_o$, $\gamma(1) = \theta_1$. Suppose $\phi_j(U)$ contains the straight line segment between $\phi_j(\theta_o)$ and $\phi_j(\theta_1)$. Then clearly

$$(6.3.4\text{-}3) \qquad d_j(\theta_o,\theta_1) = \min_{\gamma\in\Gamma(\theta_o,\theta_1,U)} I_j(\gamma).$$

The length of a curve $\gamma \in \Gamma(\theta_o,\theta_1,U)$ with respect to the Riemannian metric with Riemannian metric tensor $R(\theta,j)$ is given by the formula

$$(6.3.4\text{-}4) \qquad I(\gamma) = \int_0^1 \left\{ \frac{d\phi_j(\gamma(t))}{dt}^T R(\gamma(t),j) \frac{d\phi_j(\gamma(t))}{dt} \right\}^{\frac{1}{2}} dt$$

For each pair of points $\theta_o,\theta_1 \in C_j$ the inner metric d (with respect to the Riemannian metric) is given by

$$(6.3.4\text{-}5) \qquad d(\theta_o,\theta_1) = \inf_{\gamma\in\Gamma(\theta_o,\theta_1,M)} I(j).$$

Recall the well-known definition of equivalence of metrics.

6.3.4-6. Definition. Two metrics d',d'' on a space S are called equivalent, if there exists a constant $K_e \in \mathbb{R}_+$ such that

$$(6.3.4\text{-}7) \qquad \forall x,y \in S: d'(x,y) \le K_e d''(x,y) \text{ and } d''(x,y) \le K_e d'(x,y).$$

6.3.4-8. Proposition. Let $\theta \in \bar{C}_j'$. There exists an open neighbourhood $N = N_{\theta,j} \subseteq C_j$, $\theta \in N$, such that d_j and d are equivalent on N.

Proof. Let $C_j^{\frac{1}{2}}$ be an open set such that

$$(6.3.4\text{-}9) \qquad \bar{C}_j' \subseteq C_j^{\frac{1}{2}} \subseteq \bar{C}_j^{\frac{1}{2}} \subseteq C_j.$$

Because $R(\theta,j)$ is a positive definite matrix depending smoothly and hence continuously on θ in C_j, it follows that on the compact set $\bar{C}_j^{\frac{1}{2}}$ this matrix has a maximum eigenvalue $\lambda_M > 0$ and a minimum eigenvalue $\lambda_m > 0$. It follows that

if γ lies entirely in $\bar{C}_j^{\frac{1}{2}}$ then

(6.3.4-10) $\lambda_m I_j(\gamma) \leq I(\gamma) \leq \lambda_M I_j(\gamma)$.

Consider a fixed $\theta \in \bar{C}_j'$ and an open neighbourhood N_1 of θ, $N_1 \subseteq C_j^{\frac{1}{2}}$, such that $\phi_j(N_1) \subseteq \mathbb{R}^d$ is convex. It follows that

(6.3.4-11) $\forall \theta_o,\theta_1 \in N_1$: $d_j(\theta_o,\theta_1) = \min_{\gamma \in \Gamma(\theta_o,\theta_1,N_1)} I_j(\gamma)$.

According to a theorem from Riemannian geometry there exists an open neighbourhood $N \subseteq N_1$, $\theta \in N$ and a number $\varepsilon > 0$, such that any two points in N can be joined by a unique geodesic of length $< \varepsilon$. This geodesic lies entirely in N_1 (cf. [Boo], chapter VII, theorem (6.9) pp. 336-337). It follows that

(6.3.4-12) $\forall \theta_o,\theta_1 \in N$: $d(\theta_o,\theta_1) = \min_{\gamma \in \Gamma(\theta_o,\theta_1,N_1)} I(\gamma)$.

It follows that for all $\theta_o,\theta_1 \in N$:

(6.3.4-13) $d_j(\theta_o,\theta_1) \leq \lambda_m^{-1} d(\theta_o,\theta_1)$

and

(6.3.4-14) $d(\theta_o,\theta_1) \leq \lambda_M d_j(\theta_o,\theta_1)$.

Taking $K_e := \max(\lambda_m^{-1},\lambda_M)$ gives the desired result.

Q.E.D.

The set $\{N_{\theta,j} | \theta \in \bar{C}_j'\}$ forms an open cover of $\bar{C}_j'$, and so there is a finite subcover, because $\bar{C}_j'$ is compact. This holds for each $j \in J$. Taking the union of the finite subcovers for the different values of $j \in J$, one obtains a finite cover of M, which we will denote by $\{E_{\beta j}\}$. Of course

(6.3.4-15) $\bar{C}_j' \subseteq \bigcup_\beta E_{\beta j}$,

and for each β and j there exists a positive constant $K_{\beta j}$ such that

(6.3.4-16) $\forall \theta_o,\theta_1 \in E_{\beta j}$: $d(\theta_o,\theta_1) \leq K_{\beta j} d_j(\theta_o,\theta_1)$ and $d_j(\theta_o,\theta_1) \leq K_{\beta j} d(\theta_o,\theta_1)$.

Take

$$K_e := \max_{\beta,j} K_{\beta j} > 0.$$

It follows that

(6.3.4-17)

$$\forall\beta,\forall j,\forall\theta_o,\theta_1 \in E_{\beta j}: d(\theta_o,\theta_1) \leq K_e d_j(\theta_o,\theta_1) \text{ and } d_j(\theta_o,\theta_1) \leq K_e d(\theta_o,\theta_1).$$

In the previous subsection a finite cover $\{E_{\alpha j}\}$ was found with the property

(6.3.4-18) $$\bar{C}'_j \subseteq \bigcup_\alpha E_{\alpha j}.$$

Now let us consider the finite cover $\{E_{\alpha j} \cap E_{\beta j}\}$ of M, which we shall denote by $\{E_{ij} | i \in I(j), j \in J\}$,

(6.3.4-19) $$\bar{C}'_j \subseteq \bigcup_{i\in I(j)} E_{ij}.$$

According to proposition (6.3.2-2) there exists an n.d.s. cover $\{(E''_{ij},E'_{ij},E_{ij}) | i \in I(j),\ j \in J\}$ of M with the property

(6.3.4-20) $$\bigcup_{i\in I(j)} E''_{ij} = C'_j.$$

Now form $\{(D''_{ij},D'_{ij},D_{ij}) | i \in I(j), j \in J\}$ as in (6.3.2-10). According to proposition (6.3.2-11) this is a second order n.d.s. cover. Summarizing this section 6.3, we have found the following result

6.3.4-21. Theorem. Let $\{(C''_j,C'_j,C_j) | j \in J\}$ be an n.d.s. cover of M consisting of coordinate charts. Let $\varepsilon \in (0,1-\lambda_o)$ be fixed. Then there exists an n.d.s. cover $\{(E''_{ij},E'_{ij},E_{ij}) | i \in I(j),\ j \in J\}$ of M with the properties

(a) $E_{ij} \subseteq C_j$ for all $i \in I(j)$, $j \in J$,

(b) $\bigcup_{i\in I(j)} E''_{ij} = C'_j$ for all $j \in J$,

(c) for each $i \in I(j)$, $j \in J$ there exists a nonsingular matrix T_{ij} such that

(6.3.4-22) $\forall\theta \in E_{ij}$: $F(\phi_j(\theta);i,j) \equiv T_{ij}F(\phi_j(\theta),j)T_{ij}^{-1}$

has spectral norm smaller than $\lambda_o + \varepsilon(< 1)$,

(d) there exists a positive number K_e such that for all $i \in I(j)$, $j \in J$,

(6.3.4-23) $\forall\theta_o,\theta_1 \in E_{ij}$: $d_j(\theta_o,\theta_1) \leq K_e d(\theta_o,\theta_1)$ and $d(\theta_o,\theta_1) \leq K_e d_j(\theta_o,\theta_1)$.

There exists a corresponding second order n.d.s. cover $\{(D''_{ij},D'_{ij},D_{ij}) | i \in I(j), j \in J\}$ with $D''_{ij} = E''_{ij}$, $D'_{ij} = E'_{ij} \cap C'_j$ and $D_{ij} = E_{ij} \cap C'_j$. A fortiori, properties (a), (b), (c), (d) hold with $(E''_{ij},E'_{ij},E_{ij})$ replaced by $(D''_{ij},D'_{ij},D_{ij})$.

6.4. On the asymptotic behaviour of the stepsize and the coordinate change times

One of the main problems in analyzing algorithms like the one under investigation is the data-dependence of the parameter sequence $\{\theta_t\}$. If one fixes the sequence of parameters $\{\theta_t\}_{t=t_o}^{\infty}$ and a corresponding sequence of coordinate chart indices $\{i_t,j_t\}_{t=t_o}^{\infty}$, $i_t \in I(j_t)$, $j_t \in J$, then the state equation of the algorithm (cf. (6.2.9-8)) becomes linear (time-varying), which makes it more tractable. From the algorithm state vector $\xi(t,j)$ one can compute $e(t,j)$ and $\Psi(t,j)$ as in (6.2.9-9) and one can compute $a_t g_1 g_2 R^{-1} h$ (compare (6.2.9-11). However, in general, the equation (6.2.9-11) no longer makes sense, because the right-hand side may be outside the image $\phi_{j_t}(C_{j_t}) \subseteq \mathbb{R}^d$, and then $\hat{\theta}_{t+1}$ is no longer well-defined by (6.2.9-11). Also, of course, the so-called coupling equation (6.2.9-12) and the rules for coordinate change that follow (6.2.9-12) no longer make sense because the parameters $\{\theta_t\}$ and the coordinate charts are fixed a priori. The set of equations that compute $\xi(t,j)$, $e(t,j)$, $\Psi(t,j)$ and $a_t g_1 g_2 R^{-1} h$ in the case of a fixed sequence $\{\theta_t,i_t,j_t\}_{t=t_o}^{\infty}$ with $\theta_t \in D(i_t,j_t)$, $i_t \in I(j_t)$ and $j_t \in J$, will be called the decoupled algorithm. If equation (6.2.9-11) happens to define a sequence $\{\hat{\theta}_t\}$, i.e. if the right-hand side of (6.2.9-11) happens to be

in $\phi_{j_t}(C_{j_t})$ for each $t \geq t_o$, then and only then, the equation (6.2.9-11) will be considered as part of the decoupled algorithm.

Our approach will be to derive results for the decoupled algorithm, in which special types of parameter/index sequences $\{\theta_t, i_t, j_t\}_{t=t_o}^{\infty}$ are used, and to show what the implications are for the 'coupled' algorithm.

In this subsection we want to analyze the stepsizes and the number of steps minimally taken in each coordinate chart. Because the asymptotic behaviour is investigated, the estimates on the stepsizes etc. do not have to be sharp, they only have to be sufficient for our purposes.

The different types of parameter sequences will be described in the form of numbered properties.

6.4-1. Property 0. There exists a sequence $\{a_t'\}_{t=-\infty}^{\infty}$ with the properties

(a) for all $t < t_o$, $a_t' = a_{t_o}'$;

(b) $\{a_t'\}$ is monotonically non-increasing;

(c) $\forall t$: $a_t' > 0$ and

(d) $\sum_{t=t_o}^{\infty} a_t' = \infty$,

such that the parameter sequence $\{\theta_t^+\}_{t=-\infty}^{\infty}$ with $\theta_t^+ = (\theta_t, i_t, j_t) \in \Theta^+$ for each t, satisfies the following:

(i) $\forall t < t_o$: $\theta_t^+ = \theta_{t_o}^+$,

(ii) $\forall t \geq t_o$: $d(\theta_{t+1}, \theta_t^o) \leq a_t'$,

(iii) a coordinate change can take place only if θ_t is close enough to the boundary of $D'_{i_t j_t}$, to be precise:

$$(i_{t+1}, j_{t+1}) \neq (i_t, j_t) \Rightarrow d(\theta_t, \partial D'_{i_t, j_t}) \leq a_t',$$

(iv) a coordinate change has to take place before θ_t is too far from $D'_{i_t j_t}$, to be precise:

$$\forall t: d(\theta_t, \bar{D}'_{i_t j_t}) \leq a_t',$$

(v) if a coordinate change takes place then θ_{t+1} has to be close enough to $\bar{D}''_{i_{t+1}, j_{t+1}}$; to be precise:

$$(i_{t+1},j_{t+1}) \neq (i_t,j_t) \Rightarrow d(\theta_{t+1},\bar{D}''_{i_{t+1}j_{t+1}}) \leq a'_t.$$

6.4-2. Notation. $\lim_{t\to\infty} a'_t =: a'_\infty$.

6.4-3. Property 1. Property 0 holds and the sequence $\{a'_t\}_{t=t_o}^{\infty}$ has the extra property $a'_\infty = 0$.

6.4-4. Notation.
Before we proceed let us first introduce a useful notation. The matrices that we encounter, like G, H, K, R^{-1} etc. are all smooth functions of the local coordinates and therefore their spectral norm takes on a maximum on each $\bar{D}'_{ij}$. Because the number of indices i,j is finite it follows that for each of the matrices there is an over-all upper bound for the norm. Without loss of generality this upper bound can and will be taken greater than or equal to one. Let it be denoted by $\bar{K}(G),\bar{K}(H)$, etc.

6.4-5. Theorem. The sequence $\{\theta^+_t\}_{t\geq t_o}$ that is produced by the (coupled) algorithm satisfies property 1.

Proof. This is mainly a consequence of the use of g_2 in the parameter update equation in the algorithm (cf. (6.2.9-11) and (6.2.8-6)). It can easily be shown that

$$\|a_t g_1 g_2 R^{-1} h\| \leq a_t K_t \bar{K}(R^{-1}). \tag{6.4-6}$$

Consider the curve γ given in local coordinates of C_j by

$$\phi_j(\gamma(\lambda)) = \phi_j(\hat{\theta}_t) + \lambda a_t g_1(v_t) g_2(t) R^{-1} h, \quad \lambda \in [0,\lambda_t].$$

Clearly, by definition of λ_t (cf. (6.2.9-1)), $\gamma(\lambda_t) = \hat{\theta}_{t+1}$, and $\forall\lambda \in [0,\lambda_t]$, $\gamma(\lambda) \in \bar{C}'_j$. It follows that the Riemannian length of γ is bounded from above by $a_t K_t \bar{K}(R^{-1})\bar{K}(R)$. A fortiori one has

$$d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t K_t \bar{K}(R^{-1})\bar{K}(R).$$

Using the coupling equation (6.2.9-12), one finds

$$d(\theta_{t+1},\theta_t) \leq d(\hat{\theta}_{t+1},\theta_{t+1}) + d(\hat{\theta}_{t+1},\hat{\theta}_t) + d(\hat{\theta}_t,\theta_t) \leq$$

$$\leq \delta_{t+1} + a_t K_t \bar{K}(R^{-1})\bar{K}(R) + \delta_t.$$

Recall that $\lim_{t\to\infty} \delta_t = 0$ and $\lim_{t\to\infty} a_t K_t = 0$. Therefore one can take

$$a_t' := \max_{s\geq t}\{\delta_{s+1} + a_s K_s \bar{K}(R^{-1})\bar{K}(R) + \delta_s\},$$

to obtain a monotonically non-increasing sequence with $a_\infty' = \lim_{t\to\infty} a_t' = 0$, $a_t' > 0$ for all t and $\Sigma a_t' = \infty$. Using this sequence $\{a_t'\}$ it follows that the parameter sequence $\{\theta_t^+\}_{t=t_o}^{\infty}$ from the (coupled) algorithm satisfies $d(\theta_{t+1},\theta_t) \leq a_t'$. Using the fact that $a_t' \geq \delta_t$ and $a_t' \geq \delta_{t+1}$, it can easily be checked that also the other conditions of property 1 are satisfied. (This is left to the reader).

Q.E.D.

The idea is now to show results for the decoupled algorithm with parameter sequences satisfying property 0 with $a_\infty' > 0$ 'small enough' or property 1. Property 1, i.e. $a_\infty' = 0$ will be the most important case but the results will be needed also for $a_\infty' > 0$ now and then. Using the previous theorem, one can then draw conclusions for the coupled algorithm.

Of course one of the main problems is to estimate the effects of the coordinate changes. It has to be demonstrated that these effects do not destabilize the algorithm. One of the main results of this subsection will be that if the sequence $\{\hat{\theta}_t\}$ is well-defined by the decoupled algorithm, then the stepsize $d(\hat{\theta}_{t+1},\hat{\theta}_t)$ is bounded by $a_t\bar{K}$ for some positive constant $\bar{K}$ which is independent of t and the data.

In the algorithm at each time instance t there is a coordinate chart D_{ij} in which the algorithm operates. Let the indices i,j at time t be denoted by i_t and j_t. Consider the following indicator functions of coordinate changes

$$\chi_2(t) = \begin{cases} 1 \text{ if } j_{t+1} \neq j_t, \\ 0 \text{ elsewhere, and} \end{cases}$$

(6.4-7)

$$\chi_1(t) = \begin{cases} 1 \text{ if } j_{t+1} = j_t \text{ and } i_{t+1} \neq i_t, \\ 0 \text{ elsewhere.} \end{cases}$$

As a first step we present a lemma which gives a sequence $\{a_t \bar{\bar{v}}_t^2\}$ of upper bounds for the stepsizes.

6.4-8. Lemma. Let $\lambda_2 \in (\lambda_o, 1)$ be fixed. Let $\{(D''_{ij}, D'_{ij}, D_{ij}) \mid i \in I(j), j \in J\}$ be as in theorem (6.3.4-21) with $0 < \varepsilon < \lambda_2 - \lambda_o$. There exist positive data-independent constants c_o, c_1, c_2 such that if a sequence $\{\bar{\bar{v}}_t\}_{t_o}^{\infty}$ is defined by

(6.4-9)
$$\text{(i)} \quad \bar{\bar{v}}_{t_o-1} = 0, \text{ and}$$
$$\text{(ii)} \quad \bar{\bar{v}}_{t+1} = \lambda_2\{1 + c_1\chi_1(t) + c_2\chi_2(t)\}\bar{\bar{v}}_t + c_o\|y_{t+1}\|, \forall t \geq t_o - 1,$$

then

(6.4-10) $\forall t \geq t_o$: $d(\hat{\theta}_{t+1}, \hat{\theta}_t) \leq a_t \bar{\bar{v}}_t^2$, if $\{\hat{\theta}_t\}$ is well-defined,

(6.4-11) $\forall t \geq t_o$: $\|g_1(v_t)R^{-1}h\|_R \leq \bar{\bar{v}}_t^2$, and

(6.4-12) $\forall t \geq t_o$: $\|\xi_t\| \leq \bar{\bar{v}}_t$.

Here the quantities are quantities of the decoupled algorithm.

6.4-13. Remarks. (i) Recall that $\xi(t_o, j) = 0$ is assumed, this is of importance here.

(ii) It is not necessary to assume in this lemma that the parameter sequence $\{\theta_t^+\}$ in the decoupled algorithm satisfies property 0 or property 1.

Proof of the lemma. The two-vector $(\chi_1(t), \chi_2(t))$ can take three values, namely (0,0), (1,0) and (0,1). Each of these cases will be treated separately.

(a) $(\chi_1(t), \chi_2(t)) = (0,0)$.

Consider the algorithm equations (cf. section 6.2.9 and section 6.3.3)

$$\xi(t+i, i, j) = F(\phi; i, j)\xi(t; i, j) + G(\phi; i, j)y_t,$$

$$z(t,j) := \begin{bmatrix} e(t,j) \\ \Psi_1(t,j) \\ \vdots \\ \Psi_d(t,j) \end{bmatrix} = H(\phi;i,j)\xi(t;i,j) + Ky_t,$$

with $\phi = \phi_j(\theta_j)$, and if $\{\hat{\theta}_t\}$ is well-defined,

$$\phi_j(\hat{\theta}_{t+1}) = \phi_j(\hat{\theta}_t) + \lambda_t a_t g_1 g_2 R^{-1} h,$$

with $\lambda_t = 1$, as no coordinate change takes place in the case considered here. Taking norms, one obtains

$$\|\xi(t+1;i,j)\| \leq \|F(\phi;i,j)\|_S \|\xi(t;i,j)\| + \|G(\phi;i,j)\|_S \|y_t\|, \tag{6.4-14}$$

$$\|z(t,j)\| \leq \|H(\phi;i,j)\|_S \|\xi(t;i,j)\| + \|K\|_S \|y_t\|, \tag{6.4-15}$$

and if $\{\hat{\theta}_t\}$ is well-defined,

$$\|\phi_j(\hat{\theta}_{t+1}) - \phi_j(\hat{\theta}_j)\| \leq a_t \|R^{-1}\|_S \|h\|. \tag{6.4-16}$$

From theorem (6.3.4-21) it follows that the spectral norm of F is bounded from above by $\lambda_o + \varepsilon < \lambda_2$ and furthermore that $d(\theta_o,\theta_1) \leq K_e d_j(\theta_o,\theta_1)$ for all $\theta_o,\theta_1 \in E_{ij}$. Applying this to (6.4-14,15,16) one obtains

$$\|\xi(t+1;i,j)\| \leq (\lambda_o+\varepsilon)\|\xi(t;i,j)\| + \bar{K}(G)\|y_t\|, \tag{6.4-17}$$

$$\|z(t,j)\| \leq \bar{K}(H)\|\xi(t;i,j)\| + \bar{K}(K)\|y_t\|, \tag{6.4-18}$$

and if $\{\hat{\theta}_t\}$ is well-defined,

$$d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq K_e\, a_t\, \bar{K}(R^{-1})\|h\|. \tag{6.4-19}$$

Because $h = \Psi(t,j)^T e(t,j)$ and $z(t,j)^T = [e(t,j)^T, \Psi_1(t,j)^T, \ldots, \Psi_d^T(t,j)]$ it follows that

$$\|h\| \leq \sum_{i=1}^{d} \|\Psi_i(t,j)^T e(t,j)\| \leq \sum_{i=1}^{d} \|\Psi_i(t,j)\|\|e(t,j)\| \leq$$

$$\leq \sum_{i=1}^{d} \|z(t,j)\|\|z(t,j)\| = d\|z(t,j)\|^2. \tag{6.4-20}$$

So

$$(6.4\text{-}21)\qquad d(\hat{\theta}_{t+1},\hat{\theta}_t) \le a_t K_e \bar{K}(R^{-1}) d \|z(t,j)\|^2 = a_t \{K_e^{\frac{1}{2}} \bar{K}(R^{-1})^{\frac{1}{2}} d^{\frac{1}{2}} \|z(t,j)\|\}^2,$$

if $\{\hat{\theta}_t\}$ is well-defined.
Also,

$$(6.4\text{-}22)\qquad \|g_1(v_t)R^{-1}h\|_R \le \bar{K}(R^{-1})^{\frac{1}{2}} \|h\| \le \bar{K}(R^{-1})^{\frac{1}{2}} d \|z(t,j)\|^2.$$

(b) $(\chi_1(t),\chi_2(t)) = (1,0)$.
In this case $j(t+1) = j(t)$ and $i(t+1) \neq i(t)$. Let $i' = i(t+1)$, $i = i(t)$. Then ξ is premultiplied by a matrix $T_{i',i}$ (representing the change of basis of the state space of the algorithm). Because there are only a finite number of indices i, i.e. $\left| \bigcup_{i \in J} I(j) \right| < \infty$, there exists a finite upperbound for the spectral norms of all the matrices $T_{i',i}$. Let $c_1' > 0$ be such that $1+c_1'$ is such an upperbound. Then

$$(6.4\text{-}23)\qquad \begin{aligned} \|\xi(t+1;i_{t+1},j_{t+1})\| &\le (1+c_1')\|\xi(t+1;i_t,j_t)\| \le \\ &\le (1+c_1')(\lambda_o+\varepsilon)\|\xi(t;i_t,j_t)\| + (1+c_1')\bar{K}(G)\|y_t\|. \end{aligned}$$

(c) $(\chi_1(t),\chi_2(t)) = (0,1)$.
In this case j_t is changed into j_{t+1}. Let $j = j_t$ and $j' = j_{t+1}$, $i = i_t$ and $i' = i_t$. In general $\xi = \xi(t+i;i_t,j_t)$ will now be premultiplied by <u>three</u> matrices, to obtain $\xi(t+i;i_{t+1},j_{t+1})$:

1) By a matrix $T_i(j)$ to transform back to the C_j-state space basis of the algorithm.
2) By a matrix $S(\theta;j',j)$, cf. (6.2.7-7).
3) By a matrix $T_{i'}(j')^{-1}$ to transform to the $D(i',j')$-state-space basis of the algorithm.

Of course, 1 and 3 can be treated analogously to (b). Because $S(\theta;j',j)$ is continuous, a constant $c_2'' > 0$ can be found such that

$$(6.4\text{-}24)\qquad \|S(\theta;j',j)\|_S \le 1+c_2'', \quad \forall \theta \in \bar{C}_{j'}'' \cap \partial C_j', \ \forall j,j' \in J.$$

Combining (1), (2) and (3) it can be concluded that a constant $c_2' > 0$ exists such that

$$\|\xi(t+1;i_{t+1},j_{t+1})\| \leq (1+c_2')\|\xi(t+1;i_t,j_t)\|. \tag{6.4-25}$$

Let $\xi_t := \xi(t;i_t,j_t)$. Combining (a), (b) and (c) one obtains

$$\|\xi_{t+1}\| \leq (\lambda_o+\varepsilon)(1+c_1'\chi_1(t) + c_2'\chi_2(t))\|\xi_t\| + (1+c_1')(1+c_2')\bar{K}(G)\|y_t\|. \tag{6.4-26}$$

First, let $u_t \geq 0$ satisfy

$$u_{t_o} = 0 \tag{6.4-27}$$

and

$$u_{t+1} = (\lambda_o+\varepsilon)(1+c_1'\chi_1(t) + c_2'\chi_2(t))u_t + (1+c_1')(1+c_2')\bar{K}(G)\|y_t\|. \tag{6.4-28}$$

Then clearly $u_t \geq \|\xi_t\|$ for all $t \geq t_o$ (recall $\xi_{t_o} = 0$).
Next consider

$$\begin{cases} \bar{u}_t := K_e^{\frac{1}{2}}\bar{K}(R^{-1})^{\frac{1}{2}}d^{\frac{1}{2}}\{\bar{K}(H)u_t + \bar{K}(K)\|y_t\|\}, \forall t \geq t_o, \\ \bar{u}_{t_o-1} := 0. \end{cases} \tag{6.4-29}$$

Then

$$\begin{aligned} \bar{u}_t &\geq K_e^{\frac{1}{2}}\bar{K}(R^{-1})^{\frac{1}{2}}d^{\frac{1}{2}}\{\bar{K}(H)\|\xi_t\| + \bar{K}(K)\|y_t\|\} \geq \\ &\geq K_e^{\frac{1}{2}}\bar{K}(R^{-1})^{\frac{1}{2}}d^{\frac{1}{2}}\|z(t,j)\|, \quad \forall t \geq t_o, \end{aligned} \tag{6.4-30}$$

so

$$d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t\bar{u}_t^2, \forall t \geq t_o, \text{ if } \{\hat{\theta}_t\} \text{ is well-defined, and} \tag{6.4-31}$$

$$\|g_1(v_t)R^{-1}h\|_R \leq \bar{u}_t^2K_e^{-1}\bar{K}(R^{-1})^{-\frac{1}{2}} \leq \bar{u}_t^2. \tag{6.4-32}$$

Now $\{\bar{u}_t\}_{t=t_o-1}^{\infty}$ can be considered as the output of a linear system with inputs $\{\|y_{t+1}\|\}_{t_o-1}^{\infty}$, and zero initial conditions, as follows:

$$(6.4\text{-}33)\qquad \begin{bmatrix} u_{t+1} \\ \|y_{t+1}\| \end{bmatrix} = \begin{bmatrix} (\lambda_o+\varepsilon).(1+c_1'\chi_1(t)+c_2'\chi_2(t)) & (1+c_1')(1+c_2')\bar{K}(G) \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} u_t \\ \|y_t\| \end{bmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \|y_{t+1}\|, \quad \forall t \geq t_o-1,$$

with $u_{t_o-1} := 0$, $y_{t_o-1} := 0$ and $\chi_1(t_o-1) = 0$, $\chi_2(t_o-1) = 0$ by convention. The output equation is (6.4-29). Let T be a nonsingular matrix such that

$$T\begin{bmatrix} \lambda_o+\varepsilon & (1+c_1')(1+c_2')\bar{K}(G) \\ 0 & 0 \end{bmatrix} T^{-1}$$

has spectral norm smaller than or equal to λ_2 (recall that $\lambda_2 > \lambda_o+\varepsilon$; such a T exists according to proposition (6.3.3-5)). Let $c_1 > 0$, $c_2 > 0$ be such that the spectral norm of

$$T\begin{bmatrix} (\lambda_o+\varepsilon).(1+c_1'\chi_1(t) + c_2'\chi_1(t)) & (1+c_1')(1+c_2')\bar{K}(G) \\ 0 & 0 \end{bmatrix} T^{-1}$$

is bounded from above by $\lambda_2(1+c_1\chi_1(t)+c_2\chi_2(t))$ for all $t \geq t_o$.
Let

$$(6.4\text{-}34)\qquad \bar{\bar{u}}_t := \|T\begin{pmatrix} u_t \\ \|y_t\| \end{pmatrix}\|, \quad \forall t \geq t_o-1,$$

then

$$(6.4\text{-}35)\qquad \bar{\bar{u}}_{t+1} \leq \lambda_2(1+c_1\chi_1(t) + c_2\chi_2(t))\bar{\bar{u}}_t + \|T\|_S\|y_{t+1}\|, \quad \forall t \geq t_o-1$$

and from (6.4-29) it follows that

$$(6.4\text{-}36)\qquad \bar{u}_t \leq K_e^{\frac{1}{2}}\bar{K}(R^{-1})^{\frac{1}{2}}d^{\frac{1}{2}}\|(\bar{K}(H),\bar{K}(K))T^{-1}\|\bar{\bar{u}}_t, \quad \forall t \geq t_o.$$

Now let

(6.4-37) $c_o := K_e^{\frac{1}{2}}\bar{K}(R^{-1})^{\frac{1}{2}}d^{\frac{1}{2}}\|(\bar{K}(H),\bar{K}(K))T^{-1}\|\|T\|_S$

and

$$\begin{cases} \bar{\bar{v}}_{t_o-1} = 0 \text{ and} \\ \bar{\bar{v}}_{t+1} = \lambda_2(1+c_1\chi_1(t) + c_2\chi_2(t))\bar{\bar{v}}_t + c_o\|y_{t+1}\|, \quad \forall t \geq t_o-1, \end{cases}$$

then $\bar{u}_t \leq \bar{\bar{v}}_t$, $\forall t \geq t_o$ and so from (6.4-31) one has

(6.4-38) $d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t\bar{\bar{v}}_t^2$, if $\{\hat{\theta}_t\}$ is well-defined,

and

(6.4-39) $\|g_1(v_t)R^{-1}h\|_R \leq \bar{\bar{v}}_t^2$.

Furthermore, combining the inequalities: $\forall t \geq t_o$: $\|\xi_t\| \leq u_t$, $u_t \leq \bar{u}_t$ (this follows from (6.4-29)) and $\bar{u}_t \leq \bar{\bar{v}}_t$, it follows that

(6.4-40) $\|\xi_t\| \leq \bar{\bar{v}}_t$, $\forall t \geq t_o$.

Q.E.D.

6.4-41. Lemma. Consider the decoupled algorithm and assume its parameter sequence $\{\theta_t^+\}$ satisfies property 0 with $a_\infty' > 0$ sufficiently small or property 1. Then

(a) there exists a data-independent positive constant c_3 such that $\forall t \geq t_o$: $\bar{\bar{v}}_t \leq c_3 v_t$, and

(b) $\forall c_1,c_2 \geq 0$, $\forall \delta > 0, \exists n_1$ such that

$$\forall t \geq t_o, \forall k \geq n_1: \prod_{\tau=t}^{t+k-1}(1+c_1\chi_1(\tau) + c_2\chi_2(\tau)) < (1+\delta)^k.$$

Proof. The idea of the proof is the following. Due to property 0 the number of times that a coordinate change takes place in some given phase of the algorithm is bounded.

The proof of (a) will be given in seven steps, of which we first give an

overview.

1. Because the parameter sequence $\{\theta_t^+\}$ satisfies property 0 with $a_\infty' > 0$ small enough or property 0, the step-length $d(\theta_{t+1},\theta_t)$ has a small nonnegative limes superior for $t \to \infty$. Define $\{k_t\}$, $k_t \uparrow k_\infty \in \mathbb{R} \cup \{\infty\}$, such that on the time interval $[t,t+k_t]$, $\chi_1 = 1$ occurs at most once and $\chi_2 = 1$ occurs at most once.

2. For t large enough, $t \geq t_1$, the product $\prod_{\tau=t}^{t+k-1} \{\lambda_2(1+c_1\chi_1+c_2\chi_2)\}$ is majorized by the expression $\{(1+c_1)(1+c_2)\}^{1-k/\ell}.\lambda_3^k$, for a certain ℓ.

3. On $[t_1,t_1+\ell) \cap \mathbb{Z}$ one has $\bar{\bar{v}}_t < \bar{v}_t \times$ constant, where $\bar{v}_t = c_o(1+c_1)(1+c_2)v_t$.

4. For a certain m, it can be shown that on $[t_1+m,t_1+m+\ell) \cap \mathbb{Z}$ the inequality $\bar{v}_t \geq \bar{\bar{v}}_t$ holds.

5. If $\bar{v}_t \geq \bar{\bar{v}}_t$ then $\bar{v}_{t+\ell} \geq \bar{\bar{v}}_{t+\ell}$.

6. Combination of 4 and 5 gives

$$\forall t \geq t_1+m\colon \bar{v}_t \geq \bar{\bar{v}}_t.$$

7. It can be concluded that for some $c_3 > 0$, $\bar{\bar{v}}_t < c_3 v_t$, for all $t \geq t_o$.

ad 1. From proposition (6.3.2-19) it follows that at least a distance $c > 0$ has to be covered between any two occurrences of $\chi_1 = 1$. A similar result holds for χ_2, as can be shown easily. For notational simplicity let us denote the minimum of the two distances (again) by c.

Let

$$(6.4\text{-}42) \qquad k_t := \max\{n \in \mathbb{N}\colon n < (c/a_t')-2\} \cup \{0\}.$$

Then $\{k_t\}$ is monotonically nondecreasing and $\lim_{t\to\infty} k_t = k_\infty \in \mathbb{R} \cup \{\infty\}$. It follows from property 0 that on the time interval $[t,t+k_t]$, $\chi_1 = 1$ can occur at most once and $\chi_2 = 1$ can occur at most once.

ad 2. Fix $\lambda_2 \in (\lambda_o,\lambda_1)$ (cf. (6.2.8-2,3)) and $\lambda_3 \in (\lambda_2,\lambda_1)$. Define the natural number ℓ by

$$(6.4\text{-}43) \qquad \ell := \min\{i \in \mathbb{N} \mid (\frac{\lambda_3}{\lambda_2})^i > \max[(1+c_1)^4,(1+c_2)^4]\}.$$

Because $\lambda_3/\lambda_2 > 1$, ℓ is well-defined. By taking $a'_\infty > 0$ small enough $k_\infty > \ell$ will hold and if $a'_\infty = 0$, then $k_\infty = \infty > \ell$.
Let

(6.4-44) $t_1 := \min\{t : k_t \geq \ell\}$. This will be well-defined.

The following inequality will be shown,

(6.4-45)

$$\forall t \geq t_1, \forall k \in \mathbb{N}: \prod_{\tau=t}^{t+k-1} \{\lambda_2(1+c_1\chi_1(\tau)+c_2\chi_2(\tau))\} < (1+c_1)^{1-k/\ell}(1+c_2)^{1-k/\ell}.\lambda_3^k.$$

Note that the right-hand side of (6.4-45) is always smaller than or equal to $(1+c_1)(1+c_2)\lambda_3^k$ and that if $k \geq \ell$, then the right-hand side is smaller than or equal to λ_3^k.
The proof of (6.4-45) is straightforward. It is mainly a matter of counting how often $\chi_1(\tau) = 1$ resp. $\chi_2(\tau) = 1$ can occur on $[t,t+k-1] \cap \mathbb{Z}$. This is no more than $1 + [k/k_t]$ times, where $[k/k_t]$ denotes the entier of k/k_t Because

$k_t \geq \ell$ for $t \geq t_1$, $[k/k_t] \leq [k/\ell]$, and so

(6.4-46)
$$\prod_{\tau=t}^{t+k-1} \lambda_2(1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) \leq$$

$$\leq \prod_{\tau=t}^{t+k-1} \sqrt{\lambda_2}(1+c_1\chi_1(\tau)) \prod_{\tau=t}^{t+k-1} \sqrt{\lambda_2}(1+c_2\chi_2(\tau)) \leq$$

$$\leq \lambda_2^{k/2}(1+c_1)^{1+k/\ell}.\lambda_2^{k/2}(1+c_2)^{1+k/\ell} =$$

$$= (1+c_1)[\lambda_2^\ell(1+c_1)^2]^{\frac{k}{2\ell}}.(1+c_2)[\lambda_2^\ell(1+c_2)^2]^{\frac{k}{2\ell}} \leq$$

$$\leq (1+c_1)\left[\frac{\lambda_3^\ell}{(1+c_1)^2}\right]^{\frac{k}{2\ell}}.(1+c_2)\left[\frac{\lambda_3^\ell}{(1+c_2)^2}\right]^{\frac{k}{2\ell}},$$

where we make use of the definition of ℓ.
This shows (6.4-45).

ad 3. Let $\bar{v}_t := c_o(1+c_1)(1+c_2)v_t, \quad \forall t \geq t_o$. Let

$$\mu := \max\Big[1, \Big\{\frac{\lambda_2}{\lambda_1}(1+c_1+c_2)\Big\}^{t_1-t_o+\ell-1}\Big].$$

Then

(6.4-47) $\quad \mu\bar{v}_t \geq \bar{\bar{v}}_t$ for all $t \in [t_1, t_1+\ell) \cap \mathbb{Z}$,

because

$$\bar{v}_t \geq \sum_{j=o}^{t-t_o} \lambda_1^j c_o \|y_{t-j}\|$$

and

$$\bar{\bar{v}}_t \leq \sum_{j=o}^{t-t_o} \{\lambda_2(1+c_1+c_2)\}^j c_o \|y_{t-j}\|.$$

ad 4. Let $m := \min\{k \in \mathbb{N} \mid (\frac{\lambda_1}{\lambda_3})^k \geq \mu \text{ and } k \geq \ell\}$; m is well-defined because $\lambda_1 > \lambda_3$. It will now be shown that

(6.4-48) $\quad \forall t \in [t_1+m, t_1+m+\ell) \cap \mathbb{Z}: \ \bar{v}_t \geq \bar{\bar{v}}_t$.

The proof of this inequality is an application of (6.4-45). For each $t \in [t_1, t_1+\ell) \cap \mathbb{Z}$, $\bar{\bar{v}}_{t+m}$ is compared with $\bar{\bar{v}}_t$ and $\bar{v}_{t+m}$ with $\bar{v}_t$; then (6.4-47) is applied:

$$\forall t \in [t_1, t_1+\ell) \cap \mathbb{Z}:$$

$$\bar{\bar{v}}_{t+m} = \Big\{\prod_{\tau=t+1}^{t+m} \lambda_2(1+c_1\chi_1(\tau) + c_2\chi_2(\tau))\Big\}\bar{\bar{v}}_t +$$

$$\sum_{j=1}^{m} \Big\{\prod_{\tau=t+j+1}^{t+m} \lambda_2(1+c_1\chi_1(\tau) + c_2\chi_2(\tau))\Big\} c_o \|y_{t+j}\| \leq$$

(apply (6.4-45)) $\leq \lambda_3^m \cdot \bar{\bar{v}}_t + \sum_{j=1}^{m} \lambda_3^{m-j}(1+c_1)(1+c_2)c_o\|y_{t+j}\| \leq$

(apply definition of m) $\leq \lambda_1^m \mu^{-1} \cdot \bar{\bar{v}}_t + \sum_{j=1}^{m} \lambda_1^{m-j}(1+c_1)(1+c_2)c_o\|y_{t+j}\| \leq$

(apply (6.4-47)) $\leq \lambda_1^m \cdot \bar{v}_t + \sum_{j=1}^{m} \lambda_1^{m-j}(1+c_1)(1+c_2)c_o\|y_{t+j}\| =$

(by definition of $\bar{v}_{t+m}$) $= \bar{v}_{t+m}$.

ad 5. It will now be shown that if for any t, one has $\bar{v}_t \geq \bar{\bar{v}}_t$, then $\bar{v}_{t+\ell} \geq \bar{\bar{v}}_{t+\ell}$ holds. The proof is again an application of (6.4-45), $\lambda_3 < \lambda_1$ is also used:

$$\bar{\bar{v}}_{t+\ell} = \prod_{\tau=t+1}^{\ell} \{\lambda_2(1+c_1\chi_1(\tau) + c_2\chi_2(\tau))\}\bar{\bar{v}}_t +$$

$$+ \sum_{j=1}^{\ell} \{ \prod_{\tau=t+j+1}^{t+\ell} \lambda_2(1+c_1\chi_1(\tau)+c_2\chi_2(\tau))\}c_o\|y_{t+j}\| \leq$$

$$\leq \lambda_1^{\ell} \cdot \bar{v}_j + \sum_{j=1}^{\ell} \lambda_1^{\ell-j-1}(1+c_1)(1+c_2)c_o\|y_{t+j}\| = \bar{v}_{t+\ell}.$$

ad 6. Combining 4 and 5 gives

(6.4-49) $\quad \forall t \geq t_1+m\colon \bar{v}_t \geq \bar{\bar{v}}_t.$

ad 7. Let $t_2 := t_1 + m$ and let

(6.4-50) $\quad \mu' := \{\max[1, \frac{\lambda_2}{\lambda_1}(1+c_1+c_2)]\}^{t_2-t_o}c_o.$

Then, just as in (6.4-47), one finds

$$\forall t \in [t_o,t_2] \cap \mathbb{Z}\colon \bar{\bar{v}}_t \leq \mu' v_t.$$

Let $c_3 := \max [\mu', c_o(1+c_1+c_2)]$. Then it follows that

(6.4-51) $\quad \forall t \geq t_o\colon \bar{\bar{v}}_t \leq c_3 v_t.$

(b) This is a direct consequence of what was shown in step 2 above. Let $\delta > 0$, $c_1 \geq 0$ and $c_2 \geq 0$ be given. Take $\lambda_3 \in (\lambda_2,\lambda_1) \cap (\lambda_2,\lambda_2(1+\delta))$. Then (6.4-45) implies:

$$\forall t \geq t_1, \quad \forall k \geq \ell, \prod_{\tau=t}^{t+k-1} (1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) < (\frac{\lambda_3}{\lambda_2})^k < (1+\delta)^k.$$

Let $n_1 \geq \ell$ be such that

$$(1+\delta)^{n_1}\left(\frac{\lambda_2}{\lambda_3}\right)^{n_1} > (1+c_1+c_2)^{t_1-t_o}.$$

Then

$$\forall t \text{ with } t_o \leq t < t_1,\ \forall k \geq n_1: \prod_{\tau=t}^{t+k-1}(1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) \leq$$
$$(1+c_1+c_2)^{t_1-t_o}\prod_{\tau=t_1}^{t_1+k-1}(1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) <$$
$$(1+c_1+c_2)^{t_1-t_o}\left(\frac{\lambda_3}{\lambda_2}\right)^k < (1+\delta)^k, \text{ and so}$$
$$\forall t \geq t_1,\ \forall k \geq n_1: \prod_{\tau=t}^{t+k-1}(1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) < (1+\delta)^k,$$

this follows directly from the inequality above.

Q.E.D.

6.4-52. Theorem. There exists a constant $\bar{K}$, independent of time t and independent of the data with the following property. If $\{\hat{\theta}_t\}$ is well-defined by the decoupled algorithm with parameter sequence $\{\theta_t^+\}$ which is such that it satisfies property 0 with $a_\infty' > 0$ small enough or property 1, then the following inequality holds

$$\forall t \geq t_o: d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t\bar{K}.$$

Proof. Suppose $\{\hat{\theta}_t\}$ is well-defined. Combining the previous two lemmas one obtains the inequality

$$\forall t \geq t_o: d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t c_3^2 v_t^2.$$

If $v_t > K'$ then $g_1(v_t) = 0$ (cf. (6.2.8-4)) and therefore $\hat{\theta}_{t+1} = \hat{\theta}_t$ (cf. (6.2.9-11)). It follows that, if $\bar{K} = c_3^2(K')^2$,

$$\forall t \geq t_o: d(\hat{\theta}_{t+1},\hat{\theta}_t) \leq a_t\bar{K}.$$

Q.E.D.

6.4-53. Corollary. Consider the coupled algorithm. There exists a positive number K^+ such that

(6.4-54) $\quad \forall t \geq t_o: d(\theta_{t+1},\theta_t) \leq a_t K^+.$

Proof. Because the coupled algorithm satisfies property 1 (cf. theorem (6.4-3)) the previous theorem applies. One has

$$d(\theta_{t+1},\theta_t) \leq d(\theta_{t+1},\hat{\theta}_{t+1}) + d(\hat{\theta}_{t+1},\hat{\theta}_t) + d(\hat{\theta}_t,\theta_t) \leq$$

$$\leq \delta_{t+1} + a_t\bar{K} + \delta_t = a_t(\delta'_{t+1} + \bar{K} + \delta'_t).$$

Because $\{\delta'_t\}_{t=t_o}^{\infty}$ converges to zero for $t \to \infty$ (by assumption, cf. section 6.2.9), the sequence $\{\bar{K}+\delta'_t+\delta'_{t+1}\}_{t=t_o}^{\infty}$ converges to $\bar{K}$. Therefore it has a maximum K^+ (say).

Q.E.D.

6.4-55. Corollary. Consider the decoupled algorithm and assume that property 0, with $a'_\infty > 0$ small enough or property 1 is satisfied. There exists a $t_3 \geq t_o$, t_3 data-independent, such that

(6.4-56) $\quad \forall t \geq t_3: g_2(t) = 1.$

Proof. It suffices to show that $\|g_1(v_t)R^{-1}h\|_R$ is bounded by a data-independent constant (cf. (6.2.8-6) and recall $\lim_{t\to\infty} K_t = \infty$). Lemma (6.4-8) states

$$\|g_1(v_t)R^{-1}h\|_R \leq \overset{=2}{v_t} \text{ for all } t \geq t_o.$$

Analogously to the proof of theorem (6.4-52) it follows that

$$\|g_1(v_t)R^{-1}h\|_R \leq \bar{K} \text{ for all } t \geq t_o.$$

Q.E.D.

6.4-57. Theorem. Consider the decoupled algorithm and assume property 0 is

satisfied with $a'_\infty > 0$ small enough or assume property 1. The following inequality holds:

$$\|\xi_t\| \leq c_3 v_t \text{ for all } t \geq t_o.$$

Proof. This follows immediately by combining the inequality (6.4-12) of lemma (6.4-8), with lemma (6.4-41).

Q.E.D.

Using the results of this subsection a smaller class of parameter sequences $\{\theta_t^+\}$ can be used. This will be the class of parameter sequences that satisfy property 2 below.
Let

(6.4-58) $$\bar{a}_t := \max_{s \geq t} a_s, \text{ and } \bar{a} = \{\bar{a}_t\}_{t=-\infty}^{\infty}.$$

Then $\lim_{t\to\infty} \bar{a}_t = 0$, $\sum_{t=t_o}^{\infty} \bar{a}_t = \infty$ and $\{\bar{a}_t\}$ is monotonically non-increasing.

6.4-59. Property 2. There exists a data-independent constant $\tilde{K}^+ > 0$ such that property 1 holds with $a'_t = \bar{a}_t \tilde{K}^+$ for all t.

6.4-60. Theorem. The sequence $\{\theta_t^+\}_{t \geq t_o}$ that is produced by the (coupled) algorithm satisfies property 2 (with $\tilde{K}^+ = K^+$ as in corollary (6.4-53)).

Proof. From corollary (6.4-53) and the fact that $a_t \leq \bar{a}_t$ for all t, it follows that

$$\forall t \geq t_o: \; d(\theta_{t+1}, \theta_t) \leq \bar{a}_t K^+.$$

So condition (ii) of property 1 with $a'_t = \bar{a}_t K^+$ is satisfied. It can easily be checked that also the other conditions are satisfied.

Q.E.D.

This theorem can be compared with theorem (6.4-3). Notice that property 2 implies property 1 and therefore the results presented under the assumption of

property 0 with $a_\infty' > 0$ small enough or of property 1, also hold under the assumption of property 2.

6.5. Spaces of interpolation curves of parameter sequences and their topologies

In the following sections use will be made of interpolation curves of parameter sequences. In this section the topological and metrical structure of several spaces of interpolation curves and parameter sequences will be treated, and their relation to the algorithm.
To be able to handle coordinate chart changes in the analysis, consider the following parameter space

$$(6.5.1) \qquad \Theta^+ := \{(\theta,i,j) \mid j \in J,\ i \in I(j),\ \theta \in \bar{D}_{ij}'\}.$$

The topology will be such that Θ^+ consists of different components

$$\bar{D}_{ij}'^{+} := \{(\theta,i,j) \mid \theta \in \bar{D}_{ij}'\}$$

(for the definition of a component, cf. e.g. [Dug], p. 111). So each $\bar{D}_{ij}'^{+}$ is open and closed in this topology. Within each set $\bar{D}_{ij}'^{+}$, the topology of $\bar{D}_{ij}'$ is used: $\{(\theta,i,j) \mid \theta \in N \subseteq \bar{D}_{ij}'\}$ is open iff N is open in the relative topology of $\bar{D}_{ij}'$. This specifies the topology of Θ^+ completely. The projection

$p: \Theta^+ \to M,\ (\theta,i,j) \longmapsto \theta$ is continuous.
Next, consider for a given constant $c > 0$, the topological space $\overset{\circ}{L}_c$ of curves in M with global Lipschitz constant c:

$$(6.5\text{-}2) \qquad \overset{\circ}{L}_c := \{X: \mathbb{R} \to M \mid \forall t,s \in \mathbb{R}: d(X(t),X(s)) \leq c|t-s|\},$$

with the compact-open topology (cf. e.g. [Dug], chapter XII).
Now consider curves $r \longmapsto (X(r),i(r),j(r)) \in \Theta^+$ which have the 'Lipschitz'-property $X \in \overset{\circ}{L}_c$. The set of all such curves provided with the compact-open topology, forms a topological space $\overset{\circ}{L}{}_c^+$.
If the curves are defined on an interval $[a,b] \subseteq \mathbb{R}$, the corresponding sets will be denoted by $\overset{\circ}{L}_c[a,b]$, resp. $\overset{\circ}{L}{}_c^+[a,b]$, and if they are defined on an interval $(-\infty,b]$ or $[a,\infty)$ the corresponding sets will be denoted by

$\overset{\circ}{L}{}^{+}_{c}(-\infty,b]$ resp. $\overset{\circ}{L}{}^{+}_{c}[a,\infty)$. If the constant c is clear from the context, the index c will be dropped. An important subset of $\overset{\circ}{L}{}^{+}_{c}$ will be the set $\overset{\circ}{L}{}^{+}_{cP}$ of all curves in $\overset{\circ}{L}{}^{+}_{c}$ that satisfy the prescription (6.3.2-21). (Similarly the notation

$\overset{\circ}{L}{}^{+}_{cP}[a,b]$, $\overset{\circ}{L}{}^{+}_{cP}(-\infty,0]$ etc. will be used).
Other important subsets of $\overset{\circ}{L}{}^{+} = \overset{\circ}{L}{}^{+}_{c}$ are the sets $\overset{\circ}{L}{}^{+}_{ij}$ of all curves in $\overset{\circ}{L}{}^{+}_{c}$ that have a fixed coordinate chart index (i,j). They are in fact also subsets of $\overset{\circ}{L}{}^{+}_{cP}$ (no coordinate change takes place and the curve remains within $\bar{D}'_{ij}$, so the prescription (6.3.2-21) is satisfied). Furthermore, it is not difficult to see that $\overset{\circ}{L}{}^{+}_{ij}$ is topologically equivalent to its natural projection $\overset{\circ}{L}_{ij}$ on $\overset{\circ}{L}$.

Similarly $\overset{\circ}{L}{}^{+}_{ij}[a,b]$ is topologically equivalent to its natural projection, $\overset{\circ}{L}_{ij}[a,b]$, on $\overset{\circ}{L}[a,b]$.
A basic idea of [Ku-C1] is to associate a parameter <u>curve</u> with each parameter sequence, in such a way that the parameter curve forms an interpolation of the parameter sequence on a <u>shrinking</u> time scale. This time scale is adapted to the stochastic approximation coefficients $\{a_t\}$; instead of time instants t,

one uses the $r = \sum_{\tau=t_o}^{t-1} a_\tau$ as 'time parameter'. We will use linear interpolation

in each local coordinate chart $\phi_j(\bar{D}'_{ij}) \subseteq \mathbb{R}^d$, in accordance with section 6.2.9(a). The linear interpolation curve in Θ^+ that is produced by the coupled algorithm will be denoted by $(\theta(r),i(r),j(r))$. Because of the shrinking time scale, one has

$$\left(\theta\left(\sum_{\tau=t_o}^{t-1} a_\tau\right), i\left(\sum_{\tau=t_o}^{t-1} a_\tau\right), j\left(\sum_{\tau=t_o}^{t-1} a_\tau\right)\right) = (\theta_t, i_t, j_t), \quad \forall t \geq t_o.$$

6.5-3. <u>Theorem</u>. The curve $(\theta(r),i(r),j(r))$, $r \geq 0$, that is produced by the (coupled) algorithm is an element of $\overset{\circ}{L}{}^{+}_{cP}[0,\infty)$, with $c = K^+$.

Proof. The curve satisfies the prescription (6.3.2-21) by construction of the algorithm (cf. also section 6.2.9 (a)). The Lipschitz property with $c = K^+$ follows from corollary (6.4-53).

Q.E.D.

This theorem gives some indication of the relevance of the space $\overset{\circ}{L}{}^{+}_{cP}[0,\infty)$ and similar spaces for our problem. Next we will present some properties of these

spaces.

6.5-4. Notation. Let $X^+ \in \overset{\circ}{L}{}^+$ resp. $\overset{\circ}{L}{}^+[0,\infty)$. Then $\forall t \geq t_o, X^+_t$ will denote the translated curve

$$X^+_t(r) := X^+\Big(\sum_{\tau=t_o}^{t-1} a_\tau + r\Big), \text{ with } r \in \mathbb{R} \text{ resp. } r \geq 0.$$

Clearly $X^+_t \in \overset{\circ}{L}{}^+$ resp. $\overset{\circ}{L}{}^+[0,\infty)$.

6.5-5. Remark. We will also use 'backward interpolation curves' of the form

$$Y^+(-r) = X^+\Big(\sum_{\tau=t_o}^{t-1} a_\tau - r\Big) = X^+_t(-r).$$

Then $Y^+(0) = (\theta_t, i_t, j_t)$; $Y^+(-a_{t-1}) = (\theta_{t-1}, i_{t-1}, j_{t-1})$,

$Y^+(-a_{t-1}-a_{t-2}) = (\theta_{t-2}, i_{t-2}, j_{t-2})$ etc.

Clearly if $X^+ \in \overset{\circ}{L}{}^+$, then $Y^+ \in \overset{\circ}{L}{}^+$.
A fundamental fact for the analysis is the following application of the Arzela-Ascoli theorem.

6.5-6. Theorem. (a) $\forall a,b \in \mathbb{R}$ with $a < b$, $\overset{\circ}{L}[a,b]$ is compact.
(b) Any subset of $\overset{\circ}{L}[a,b]$ is relatively compact. (i.e. has compact closure).

Proof. The equicontinuity of $\overset{\circ}{L}[a,b]$ and therefore of any subset follows easily from the Lipschitz property. Therefore, according to the theorem of Arzela-Ascoli, $\overset{\circ}{L}[a,b]$ and any subset of $\overset{\circ}{L}[a,b]$ is relatively compact in $C^o([a,b],M)$, the set of all continuous curves $X: [a,b] \to M$. It now suffices to show that $\overset{\circ}{L}[a,b]$ is closed in $C^o([a,b],M)$. This is not difficult, and is left to the reader. Q.E.D.

6.5-7. Theorem (a) Any connected component of $\overset{\circ}{L}{}^+[a,b]$ is compact (and therefore each subset of such a component is relatively compact).
(b) For each index (i,j), $\overset{\circ}{L}{}^+_{ij}[a,b] \cong \overset{\circ}{L}_{ij}[a,b]$ is compact.

Proof. (a) With any curve $X^+ \in \overset{\circ}{L}[a,b]$, $X^+(r) = (X(r), i(r), j(r))$ is associated the 'index curve' $(i(r), j(r))$, $r \in [a,b]$. If any two curves $X^+, \tilde{X}^+ \in \overset{\circ}{L}{}^+[a,b]$

have different index function then they lie in different components. This can be seen as follows. Suppose $r_o \in [a,b]$ is such that $i(r_o) = i_o \neq \tilde{i}(r_o)$. The sets

$$\{Z^+ = (Z(r),i(r),j(r)) \in \overset{\circ}{L}{}^+[a,b] \,|\, i(r_o) = i_o\}$$

and

$$\{Z^+ = (Z(r),i(r),j(r)) \in \overset{\circ}{L}{}^+[a,b] \,|\, i(r_o) \neq i_o\}$$

are both open in $\overset{\circ}{L}{}^+[a,b]$, and their union is $\overset{\circ}{L}{}^+[a,b]$; X^+ is element of the first set and $\tilde{X}^+$ is element of the second set.
Because of this, it suffices to show that each subset of all curves $X^+(r)$ with a <u>fixed index curve</u> $(i(r),j(r))$ is compact. Such a subset is given by

$$\{(X,i,j) \in \overset{\circ}{L}{}^+[a,b] \,|\, X \in \overset{\circ}{L}[a,b],\ \forall r \in [a,b]\colon X(r) \in \bar{D}_{i(r)j(r)}\}.$$

This set is homeomorphic to

$$\{X \in \overset{\circ}{L}[a,b] \,|\, \forall r \in [a,b]\colon X(r) \in \bar{D}_{i(r)j(r)}\}$$

which is a closed subset of $\overset{\circ}{L}[a,b]$. (It is easily seen that the complement of this set is open). Because $\overset{\circ}{L}[a,b]$ is compact, the same holds for any closed subset.
(b) A special case of a fixed index curve is a constant index curve $(i(r),j(r)) = (i,j)$. Using this, the result follows directly from the proof of (a). Q.E.D.

Let $t \geq t_o$, $t \in \mathbb{Z}$. Then to each curve $Y^+ \in \overset{\circ}{L}{}^+(-\infty,0]$ corresponds a parameter sequence $\{(\theta_{t-\tau},i_{t-\tau},j_{t-\tau})\}_{\tau=0}^{t-t_o}$ according to the ('backward') formula

$$(6.5\text{-}8) \qquad (\theta_{t-\tau},i_{t-\tau},j_{t-\tau}) = \begin{cases} Y^+(0) & \text{if } \tau = 0, \\ Y^+\left(-\sum_{j=t-\tau}^{t-1} a_j\right), & \text{if } \tau = 1,2,\ldots,t-t_o \end{cases}$$

(cf. remark (6.5-5)). Let $\Pi_t\colon \overset{\circ}{L}{}^+(-\infty,0] \to (\theta^+)^{t-t_o+1}$ be the mapping which maps

$$Y^+ \in \overset{\circ}{L}{}^+(-\infty,0] \text{ to } \{(\theta_{t-\tau}, i_{t-\tau}, j_{t-\tau})\}_{\tau=0}^{t-t_o}$$

given by (6.5-8).

Let us formally extend the definition of the sequence $\{a_t\}$ to all $t < t_o$ by taking $a_t = a_{t_o}$ for all $t < t_o$. Let $a = \{a_t\}_{t=-\infty}^{\infty}$ denote the resulting sequence. Then Π_t^a can be defined as:

(6.5-9)
$$\Pi_t^a: \overset{\circ}{L}{}^+(-\infty,0] \to (\Theta^+)^{\mathbb{N}}$$
$$Y^+ \longmapsto \{\theta_{t-\tau}, i_{t-\tau}, j_{t-\tau}\}_{\tau=0}^{\infty} \text{ with}$$
$$(\theta_{t-\tau}, i_{t-\tau}, j_{t-\tau}) = \begin{cases} Y^+(0) & \text{if } \tau = 0, \\ Y^+\left(-\sum_{j=t-\tau}^{t-1} a_j\right), & \text{if } \tau = 1,2\ldots \end{cases}$$

6.5-10. Remarks. (i) The sets $(\Theta^+)^{t-t_o+1}$ and $(\Theta^+)^{\mathbb{N}}$ will be considered as topological spaces with the topology of pointwise convergence. This coincides with the compact-open topology, of course.

(ii) In Π_t^a, the sequence $a = \{a_t\}$ can be replaced by any sequence $\tilde{a} = \{\tilde{a}_t\}$ with $\tilde{a}_t \geq 0$, $\forall t$. Notation: $\Pi_t^{\tilde{a}}$.

It is not difficult to show:

6.5-11. Theorem. For each t and each $\tilde{a} = \{\tilde{a}_t\}$ with $\tilde{a}_t \geq 0$ for all t, $\Pi_t^{\tilde{a}}$ is continuous.

The topological spaces $\overset{\circ}{L}(-\infty,0]$, $\overset{\circ}{L}{}^+(-\infty,0]$, $\Theta^{\mathbb{N}}$ and $(\Theta^+)^{\mathbb{N}}$ can be provided with a metric in the following manner.

6.5-12. Definition. Let d denote the metric on M.

(a) Let $d^+: \Theta^+ \times \Theta^+ \to [0,\infty]$ be the mapping given by

$$d^+[(\theta,i,j),(\tilde{\theta},\tilde{i},\tilde{j})] = \begin{cases} d(\theta,\tilde{\theta}) & \text{if } i = \tilde{i} \text{ and } j = \tilde{j}, \\ \infty & \text{if } i \neq \tilde{i} \text{ or } j \neq \tilde{j}. \end{cases}$$

(b) Let $\rho_c: \overset{\circ}{L}(-\infty,0] \times \overset{\circ}{L}(-\infty,0] \to [0,1]$ be given by

$$\rho_c(X,\tilde{X}) = \sup_{n\in\mathbb{N}}[\min\{\tfrac{1}{n}, \sup_{r\in[-n,0]} d(X(r),\tilde{X}(r))\}].$$

(c) Let $\rho_c^+: \overset{\circ}{L}^+(-\infty,0] \times \overset{\circ}{L}^+(-\infty,0] \to [0,1]$ be given by

$$\rho_c^+(X^+,\tilde{X}^+) = \sup_{n\in\mathbb{N}}[\min\{\tfrac{1}{n}, \sup_{r\in[-n,o]} d^+(X^+(r),\tilde{X}^+(r))\}].$$

(d) Let $\rho: M^{\mathbb{N}_o} \times M^{\mathbb{N}_o} \to [0,1]$ be given by

$$\rho(\alpha,\beta) = \sup_{n\geq o}[\min\{\tfrac{1}{n+1}, \sup_{o\leq i\leq n} d(\alpha_i,\beta_i)\}].$$

(e) Let $\rho^+: (\Theta^+)^{\mathbb{N}_o} \times (\Theta^+)^{\mathbb{N}_o} \to [0,1]$ be given by

$$\rho(\alpha^+,\beta^+) = \sup_{n\geq 0}[\min\{\tfrac{1}{n+1}, \sup_{0\leq i\leq n} d^+(\alpha_i^+,\beta_i^+)\}].$$

6.5-13. Theorem. $\rho_c,\rho_c^+,\rho,\rho^+$ are metrics of $\overset{\circ}{L}(-\infty,0]$, $\overset{\circ}{L}^+(-\infty,0]$, M and Θ^+ respectively, and these metrics are compatible with the given topologies of these spaces (i.e. the corresponding metric topology coincides with the original topology of each space).

Proof. For ρ_c, cf. [Dug], chapter XII, p. 272, 8.5. The other cases are straightforward modifications of the case with ρ_c. This is left to the reader. Q.E.D.

Using these metrics, theorem (6.5-11) can be strengthened:

6.5-14. Proposition. Let $\alpha \geq \max\{1,\max_{t\in\mathbb{Z}} \tilde{a}_t\}$. Then for each $t \in \mathbb{Z}$:

$$\forall X^+,\tilde{X}^+ \in \overset{\circ}{L}^+(-\infty,0]: \rho(\Pi_t^{\tilde{a}}(X^+), \Pi_t^{\tilde{a}}(\tilde{X}^+)) \leq \alpha.\rho_c(X^+,\tilde{X}^+).$$

Proof. It suffices to show the inequality for all $\alpha > \max\{1,\sup_{t\in\mathbb{Z}} \tilde{a}_t\}$. One has

$$\rho_c(X^+,\tilde{X}^+) = \sup_{n\in\mathbb{N}}[\min\{\tfrac{1}{n}, \sup_{-n\leq r\leq 0} d^+(X^+(r),\tilde{X}^+(r))\}] \geq$$

$$\geq \sup_{n\in\mathbb{N}}\Big[\min\Big\{1,\ \operatorname*{supremum}_{\{\tau\geq 0 \mid \sum_{j=t-\tau}^{t-1}\tilde{a}_j\in[0,n]\}} d^+\Big(X^+\Big(-\sum_{j=t-\tau}^{t-1}\tilde{a}_j\Big),\tilde{X}^+\Big(-\sum_{j=t-\tau}^{t-1}\tilde{a}_j\Big)\Big)\Big\}\Big].$$

To simplify the notation let

$$\theta^+_{t-\tau} := X^+\Big(-\sum_{j=t-\tau}^{t-1}\tilde{a}_j\Big) \in \Theta^+$$

and

$$\tilde{\theta}^+_{t-\tau} := \tilde{X}^+\Big(-\sum_{j=t-\tau}^{t-1}\tilde{a}_j\Big) \in \Theta^+.$$

It follows that

$$\rho_c(X^+,\tilde{X}^+) \geq \sup_{n\in\mathbb{N}}\Big[\min\Big\{\frac{1}{n},\ \sup_{0\leq\alpha\tau\leq n} d^+(\theta^+_{t-\tau},\tilde{\theta}^+_{t-\tau})\Big\}\Big] \geq$$

$$\geq \frac{1}{\alpha}\sup_{n\in\mathbb{N}}\Big[\min\Big\{\frac{1}{[\frac{n}{\alpha}]+1},\ \sup_{0\leq\tau\leq[\frac{n}{\alpha}]} d^+(\theta^+_{t-\tau},\tilde{\theta}^+_{t-\tau})\Big\}\Big],$$

where as before $[\frac{n}{\alpha}]$ denotes the entier of $\frac{n}{\alpha}$. This last expression is equal to

$$\frac{1}{\alpha}\sup_{n\in\mathbb{N}}\Big[\min\Big\{\frac{1}{n},\ \sup_{0\leq\tau\leq n-1} d^+(\theta^+_{t-\tau},\tilde{\theta}^+_{t-\tau})\Big\}\Big] = \frac{1}{\alpha}\rho\{\Pi_t(X^+,\tilde{X}^+)\}.$$

Q.E.D.

6.5-15. Remark. Because $\forall t < t_o$: $a'_t = a_t$ by definition, and because $\lim_{t\to\infty} a'_t = a'_\infty \in \mathbb{R}$, $\{a'_t\}_{t=-\infty}^{\infty}$ has a finite maximum. So the proposition can be applied to $\{a'_t\}$.

The connection between some of the spaces of interpolation curves treated here and the classes of parameter sequences that were presented in the previous section will be treated next.

The mapping $\Pi^{\tilde{a}}_t$, defined earlier on the space $\overset{\circ}{L}{}^+_c(-\infty,0]$, can also be defined on $\overset{\circ}{L}{}^+_c$, as follows. Let $\Pi^{\tilde{a}}_t$ be the mapping which maps $X^+ \in \overset{\circ}{L}{}^+_c$ to $\{\theta^+_{t-\tau}\}_{\tau=-\infty}^{\infty}$ given by (compare (6.5-8))

$$\theta^+_{t+\tau} = \begin{cases} X^+\Big(\sum_{j=t}^{t+\tau-1} \tilde{a}_j\Big) & \text{if } \tau > 0, \\ X^+(0) & \text{if } \tau = 0, \\ X^+\Big(-\sum_{j=t+\tau}^{t-1} \tilde{a}_j\Big) & \text{if } t_o - t \leq \tau < 0, \text{ and} \\ X^+\Big(-\sum_{j=t_o}^{t-1} \tilde{a}_j\Big) & \text{if } \tau < t_o - t. \end{cases} \tag{6.5-16}$$

Notice that, because of the shift-invariance of $\mathring{L}^+_c$, the image $\Pi_t(\mathring{L}^+_c)$ is the same for each $t \in \mathbb{Z}$. A similar conclusion holds for any shift-invariant subspace of $\mathring{L}^+_c$ like $\mathring{L}^+_{cP}$.

6.5-17. Property 3. The parameter sequence $\{\theta^+_t\}_{t=-\infty}^{\infty}$ is an element of $\Pi^a_t(\mathring{L}^+_{cP})$.

6.5-18. Theorem. Property 3 implies property 2.

Proof. Let $\bar{a}$ be as in (6.4-58). Let $X^+ \in \mathring{L}^+_{cP}$ be such that $\{\theta^+_s\}_{s=-\infty}^{\infty} = \Pi^{\bar{a}}_t(X^+)$. Using (6.5-16) and the Lipschitz property of X one finds that for $s > t$

$$d(\theta_{s+1}, \theta_s) = d\Big(X\Big(\sum_{j=t}^{s} a_j\Big),\ X\Big(\sum_{j=t}^{s-1} a_j\Big)\Big) \leq ca_s \leq c\bar{a}_s.$$

A similar formula holds for $s \leq t$. This shows that $\{\theta^+_s\}$ satisfies condition (ii) of property 1 with $a'_s = c\bar{a}_s$. Condition (i) of property 1 follows directly from the definition of Π^a_t. A coordinate change in the parameter sequence takes place from θ^+_s to θ^+_{s+1} iff a coordinate change in X^+ takes place on the interval from $\sum_{j=t}^{s-1} a_j$ to $\sum_{j=t}^{s} a_j$.

The interval has length $a_s \leq \bar{a}_s$ and because of the Lipschitz property of X, the points θ_s and θ_{s+1} lie at a distance less than $c\bar{a}_s$ $(= a'_s)$ from the point on the curve X at which the coordinate change takes place. Because $X^+ \in \mathring{L}^+_{cP}$ it satisfies the prescription (6.3.2-21). This implies that conditions (iii), (iv) and (v) of property 1 hold for $\{\theta^+_t\}$ with $a'_s = c\bar{a}_s, \forall s$.

Q.E.D.

The following theorem can be compared to theorem (6.4-60).

6.5-19. Theorem. The sequence $\{\theta_t^+\}_{t\geq t_o}$ that is produced by the coupled algorithm satisfies property 3 (with $c = K^+$ as in corollary (6.4-53)).

Proof. This follows from theorem (6.5-3), using $a_t \leq \bar{a}_t$.

Q.E.D.

6.6. On a problem of P-a.s. convergence

6.6.1. Introduction

In this introduction we try to explain the main ideas concerning the relations between P-a.s. properties in the decoupled case and the coupled case by way of some illustrative propositions.

6.6.1-1. Example. Consider a probability space $(\Omega,\overset{\circ}{H},P)$ and a function

$$F: \Gamma \times \Omega \to \Phi$$
$$(\gamma,\omega) \longmapsto F(\gamma,\omega),$$

with Γ and Φ topological spaces, Γ separable (i.e. Γ is a Hausdorff space which contains a countable dense set). Then

6.6.1-2. Proposition. Suppose that for each $\omega \in \Omega$, F is continuous as a function of $\gamma \in \Gamma$. Let $S \subseteq \Phi$ be a closed set such that for each $\gamma \in \Gamma$:

(6.6.1-3) $\quad F(\gamma,\omega) \in S$, P-a.s.

Let $G: \Omega \to \Gamma$ be arbitrary. Then

(6.6.1-4) $\quad F(G(\omega),\omega) \in S$, P-a.s.

Proof. Let $\{\gamma_1,\gamma_2,\ldots\}$ be a countable, dense subset of Γ. Then for each $i \in \mathbb{N}$, $F(\gamma_i,\omega) \in S$, P-a.s. Therefore, for each $i \in \mathbb{N}$ there exists a set $E_i \subseteq \Omega$ of 'exceptions', with $P(E_i) = 0$ and

$$\forall \omega \in \Omega \backslash E_i: F(\gamma_i,\omega) \in S.$$

Let $E := \bigcup_{i \in \mathbb{N}} E_i$. Then $P(E) = 0$ and

$$\forall i \in \mathbb{N}: \forall \omega \in \Omega \backslash E: F(\gamma_i, \omega) \in S.$$

Now let γ be arbitrary. Because $\{\gamma_i\}$ is dense in Γ, there exists a subsequence $\{\gamma_{i(k)}\}_{k \in \mathbb{N}}$, that converges to γ. It follows that

$$\forall \omega \in \Omega \backslash E: F(\gamma, \omega) = \lim_{k \to \infty} F(\gamma_{n(k)}, \omega) \in S,$$

because S is closed. So

(6.6.1-5) $\forall \gamma \in \Gamma: \forall \omega \in \Omega \backslash E: F(\gamma, \omega) \in S.$

This implies clearly

$$\forall \omega \in \Omega \backslash E: F(G(\omega), \omega) \in S,$$

and so

$$F(G(\omega), \omega) \in S, \ P\text{-a.s.}$$

Q.E.D.

This theorem is meant to give an idea of the kind of reasoning that is involved; it will not be used. In our application of this kind of idea, γ will be a sequence or curve in the parameter space and F will be a sequence, while ω represents the data $\{y_k\}$.

Now consider the following situation. Let $\{F_k(\gamma, \omega)\}_{k=1}^{\infty}$ be a sequence of functions

(6.6.1-6) $F_k: \Gamma \times \Omega \to \Phi,$

where Γ is now a compact topological space, Φ is a normed linear space with norm $\|,\|$. (Ω, H, P) is a probability space, like in the example above. Consider the following set of functions of γ.

(6.6.1-7) $\{F_k(.;\omega) | k \in \mathbb{N}, \omega \in \Omega\}$.

Suppose it is <u>equicontinuous</u> on Γ, i.e. for each $\gamma_o \in \Gamma$ and for each $\varepsilon > 0$ there exists a neighbourhood N of γ_o such that

$$\forall k \in N, \forall \omega \in \Omega, \ \forall \gamma \in N: \ \|F_k(\gamma,\omega) - F_k(\gamma_o,\omega)\| < \varepsilon.$$

(For the general definition of equicontinuity, cf.e.g. [Dug]). Then the following holds.

6.6.1-8. <u>Proposition</u>. Suppose that for each $\gamma \in \Gamma$,

$$\lim_{k\to\infty} F_k(\gamma;\omega) = 0, \ P\text{-a.s.}$$

(a) Then $\exists N \subseteq \Omega$, with $P(N) = 0$ such that

$$\forall \omega \in \Omega\, N, \ \forall \{\gamma_k | \gamma_k \in \Gamma\}_{k=1}^{\infty}: \lim_{k\to\infty} F_k(\gamma_k,\omega) = 0.$$

(b) Let $\{G_k\}_{k\in\mathbb{N}}$ be a sequence of (arbitrary) mappings $G_k: \Omega \to \Phi$. Then

$$\lim_{k\to\infty} F_k(G_k(\omega);\omega) = 0, \ P\text{-a.s.}$$

Proof. (b) follows directly from (a), so it suffices to prove (a).
It will be shown that

(6.6.1-9) $\forall \varepsilon > 0$: $\exists N \subseteq \Omega$ such that $\forall \omega \in \Omega\backslash N$,

for all sequences $\{\gamma_k | \gamma_k \in \Gamma, k \in \mathbb{N}\}$: $\limsup_{k\to\infty} \|F_k(\gamma_k,\omega)\| \leq \varepsilon$.

Once this is established the result follows easily: Let $E_\varepsilon \subseteq \Omega$ denote the exceptions set for a given $\varepsilon > 0$, then $P(E_\varepsilon) = 0$. Let

$$E = \bigcup_{n\in\mathbb{N}} E_{1/n},$$

then $P(E) = 0$ and

$$\forall \omega \in \Omega\backslash E: \ \forall \{\gamma_k | \gamma_k \in \Gamma\}_{k=1}^{\infty}: \lim_{k\to\infty} F_k(\gamma_k,\omega) = 0.$$

Now (6.6.1-9) will be shown. Let $\varepsilon > 0$ be fixed. Because the set of functions in (6.6.1-7) is equicontinuous, for each $\gamma_o \in \Gamma$ an open neighbourhood N_{γ_o} can be found such that

(6.6.1-10) $\forall k \in \mathbb{N}, \forall \omega \in \Omega, \forall \gamma \in N_{\gamma_o}: \|F_k(\gamma,\omega) - F_k(\gamma_o,\omega)\| < \frac{\varepsilon}{2}$.

Clearly $\{N_{\gamma_o} | \gamma_o \in \Gamma\}$ forms an open cover of Γ. Because Γ is compact there exists a finite subcover, say $N_{\gamma^1}, N_{\gamma^2}, \ldots, N_{\gamma^n}$. For each $i \in \{1,2,\ldots,n\}$

(6.6.1-11) $\lim_{k\to\infty} F_k(\gamma^i,\omega) = 0$, P-a.s.

Let for each $i \in \{1,2,\ldots,n\}$, E_i be the exceptions set, $P(E_i) = 0$. Let

$$\bar{E} = \bigcup_{i=1}^{n} E_i,$$

then $P(\bar{E}) = 0$. For each $\omega \in \Omega\backslash\bar{E}$ and for each $i \in \{1,2,\ldots,n\}$ there exists a $k_i(\omega)$ such that

(6.6.1-12) $\forall k \geq k_i(\omega): \|F_k(\gamma^i,\omega)\| < \varepsilon/2$.

Let $k_o(\omega) := \max_{1\leq i\leq n} k_i(\omega)$. Then one has for each $\omega \in \Omega\backslash\bar{E}$:

(6.6.1-13) $\forall k \geq k_o(\omega), \forall i \in \{1,2,\ldots,n\}: \|F_k(\gamma^i,\omega)\| < \varepsilon/2$.

Let $\{\gamma_k\}_1^\infty$ be an arbitrary sequence in Γ. Because $\bigcup_i N_{\gamma^i} = \Gamma$, for each $k \in \mathbb{N}$ there exists an $i(k) \in \{1,2,\ldots,n\}$ such that $\gamma_k \in N_{\gamma^{i(k)}}$. Therefore applying (6.6.1-10) and (6.6.1-13),

$$\forall \omega \in \Omega\backslash\bar{E}, \ \forall k \geq k_o(\omega):$$

$$\|F_k(\gamma_k,\omega)\| \leq \|F_k(\gamma_k,\omega) - F_k(\gamma^{i(k)},\omega)\| + \|F_k(\gamma^{i(k)},\omega)\| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

So

$$P(\bar{E}) = 0 \text{ and } \forall \omega \in \Omega\backslash\bar{E}, \ \forall \{\gamma_k | \gamma_k \in \Gamma\}_{k=1}^\infty: \ \limsup_{k\to\infty} \|F_k(\gamma_k,\omega)\| \leq \varepsilon,$$

i.e. (6.6.1-9) is shown.

Q.E.D.

6.6.2. On (equi-)continuity of some variables in the algorithm

To be able to apply the ideas of section 6.6.1 in the analysis of the algorithm, one needs to establish equicontinuity of several variables in the algorithm. More precisely each of these variables will be considered to be a family of functions. The family of functions is obtained by considering the decoupled algorithm, and considering the variable involved as a function of a parameter interpolation curve, for each time t and each data sequence ω. By letting t and ω vary over all possibilities one obtains a family of functions.

6.6.2-1 Notation. Let $\omega = \{y_t\}_{t=t_o}^{\infty}$ denote the data sequence, and let

$$\forall t \geq t_o: b_t = b_{t\omega} := g_1(v_t)g_2(t)R(\theta_t, j_t)^{-1}h$$

(compare (6.2.8-5)).

So if $\{\hat{\theta}_t\}$ is well-defined and $\lambda_t a_t \neq 0$,

$$\forall t \geq t_o: b_{t,\omega} = (\lambda_t a_t)^{-1}(\phi(\hat{\theta}_{t+1}, j_t) - \phi(\hat{\theta}_t, j_t))$$

(cf. also (6.2.9-2)).

For each t and ω, $b_{t\omega}$ can be considered in the decoupled algorithm as a function $b_{t\omega}(\alpha)$ of the sequence of past parameter values

$$\alpha = \{\theta^+_{t-\tau} \in \Theta^+\}_{\tau=o}^{\infty}.$$

Now if $\{\theta^+_s\}$ satisfies property 3 then

(6.6.2-2) $$\{\theta^+_{t-\tau}\}_{\tau=o}^{\infty} \in \Pi_t^a(\mathring{L}^+_{cP}(-\infty, 0]).$$

So $b_{t\omega} \circ \Pi_t^a$ is a function of the elements of $\mathring{L}^+_{cP}(-\infty, 0]$ for each $t \geq t_o$ and each ω. To show equicontinuity of the $b_{t\omega} \circ \Pi_t^a$ the following technical lemma is needed.

6.6.2-3. Lemma. Let t_3 be as in corollary (6.4-55). For each $\varepsilon' > 0$ there exists a $\delta > 0$ such that for each $\omega \in \Omega$, for all $t \geq t_3$ and for all

$\alpha,\beta \in \Pi_t^a(\mathring{L}^+_{cP}(-\infty,0]) \subseteq (\Theta^+)^{N_o}$ the following implication holds:

$$\rho^+(\alpha,b) < \delta \Rightarrow \|b_{t\omega}(\alpha)-b_{t\omega}(\beta)\| < \varepsilon'.$$

Proof. In fact the following will be shown.

$$\forall \varepsilon' > 0, \exists \delta > 0 \text{ such that } \forall \omega \in \Omega, \forall t \geq t_3,$$
$$\forall \alpha,\beta \in \tilde{\Pi}_t^a(\mathring{L}^+_{cP}(-\infty,0]): (\rho^+(\alpha,\beta) < \delta \Rightarrow \|b_{t\omega}(\alpha)-b_{t\omega}(\beta)\| < \varepsilon')$$

where $\tilde{a} = (\ldots \bar{a}_{t_o}, \bar{a}_{t_o}, \bar{a}_{t_o}, \bar{a}_{t_o+1}, \bar{a}_{t_o+2}, \ldots, \bar{a}_{t_4-1}, \bar{a}_{t_4}, \bar{a}_{t_4} \ldots)$ with t_4 sufficiently large such that $\bar{a}_{t_4} > 0$ is sufficiently small to be able to apply the results derived before under the assumption of property 0 with $a_\infty' > 0$ sufficiently small.

The proof consists of three parts. Part (i) shows that 'what happened long ago does not have much effect' and part (ii) shows the continuity with respect to the 'recent past'. Part (iii) combines (i) and (ii). Let $\varepsilon' > 0$ be fixed. Choose $\lambda \in (\lambda_o,\lambda_1)$ (compare (6.2.8-3)), and let $\{(D''_{ij},D'_{ij},D_{ij})\}$ be as in theorem (6.3.4-21) with $\varepsilon = \lambda - \lambda_o$.

Define δ' and δ'' as follows

$$\delta' := \varepsilon' 2^{-4} d^{-1} \bar{K}(H)^{-2}(1+c_3)^{-1}(K')^{-1}(\bar{K}(R^{-1}))^{-1}, \tag{6.6.2-4}$$

$$\delta'' = (\varepsilon')^{\frac{1}{2}} 2^{-\frac{3}{2}} d^{-\frac{1}{2}} \bar{K}(H)^{-1} \bar{K}(R^{-1})^{-\frac{1}{2}}. \tag{6.6.2-5}$$

According to lemma (6.4-41) there exists for $\delta = (\frac{\lambda_1}{\lambda})^{\frac{1}{2}} - 1$ (and $c_1 \geq 0$, $c_2 \geq 0$ as in the proof of lemma (6.4-8)) a number $n_1 \in \mathbb{N}$ such that

$$\forall t \geq t_o, \forall k \geq n_1: \prod_{\tau=t}^{t+k-1}(1+c_1\chi_1(\tau)+c_2\chi_2(\tau)) < (1+\delta)^k = (\frac{\lambda_1}{\lambda})^{k/2},$$

and so

(6.6.2-6) $\forall t \geq t_o$, $\forall k \geq n_1$: $\prod_{\tau=t}^{t+k-1} \{(\frac{\lambda}{\lambda_1})(1+c_1\chi_1(\tau)+c_2\chi_2(\tau))\} < (\frac{\lambda}{\lambda_1})^{k/2}$.

Now let n_o with $n_o \geq n_1$ and $n_o \geq (ca_{t_4})^{-1}$ be such that

(6.6.2-7) $(\frac{\lambda}{\lambda_1})^{n_o/2} < \frac{\min(\delta', \delta'')}{2c_3\bar{K}}$.

This is possible because $0 < \lambda < \lambda_1$. Here c_3 is as in lemma (6.4-41).
(i) The following will be shown. Let $t \geq t_3$ be fixed. If the first $n_o + 1$ components of $\alpha,\beta \in \tilde{\Pi}_t^a(\mathring{L}_{cP}^+(-\infty,0])$ are equal, i.e. $\alpha_i = \beta_i$, $i = 0,1,2,\ldots,n_o$,

then $\|b_{t\omega}(\alpha)-b_{t\omega}(\beta)\| < \varepsilon'/4$.

Of course, if for some ω, $v_t > K'$ then $g_1(v_t) = 0$, hence $b_{t\omega}(\alpha) = 0$ and so $\|b_{t\omega}(\alpha) - b_{t\omega}(\beta)\| = 0$ in this case. Therefore for this case the assertion is correct. Now suppose ω is such that $v_t \leq K'$. It then follows easily from the definition of v_t that $\forall\tau \geq 0$: $v_{t-\tau} \leq K'\lambda_1^{-\tau}$. Because $\alpha,\beta \in \tilde{\Pi}_t^a(\mathring{L}_{cP}^+(-\infty,0])$, one can assume without loss of generality that property 0 is satisfied with $a_\infty' > 0$ small enough or property 1 is satisfied, so theorem (6.4-57) is applicable. One obtains

(6.6.2-8) $\forall\tau \geq 0$: $\|\xi_{t-\tau}\| \leq c_3K'\lambda_1^{-\tau}$,

(where by convention, one takes $\xi_t = 0$ if $t < t_o$).
Now consider the difference

$$\Delta\xi_{t-\tau} = \xi_{t-\tau}(\alpha) - \xi_{t-\tau}(\beta),$$

in an obvious notation. Now notice that if $\tau \leq n_o$ then, because $\alpha_\tau = \beta_\tau$, $\Delta\xi_{t-\tau}$ depends only on $\Delta\xi_{t-n_o}$ and on the $\alpha_\tau = \beta_\tau$, $\tau = 0,\ldots,n_o$, and not on the data $y_{t-\tau}$, $\tau = 0,\ldots,n_o$. In a manner, completely analogous to the derivation of equation (6.4-26) in the proof of lemma (6.4-8), one can derive

(6.6.2-9) $\forall\tau \in \{1,2,\ldots,n_o\}$: $\|\Delta\xi_{t-\tau+1}\| \leq \lambda(1+c_1\chi_1(t-\tau) + c_2\chi_2(t-\tau))\|\Delta\xi_{t-\tau}\|$

and therefore

(6.6.2-10) $\|\Delta\xi_t\| \leq \prod_{\tau=1}^{n_o} \lambda(1+c_1\chi_1(t-\tau)+c_2\chi_2(t-\tau))\|\Delta\xi_{t-n_o}\|$.

If $t \leq n_o$ then $\xi_{t-n_o} = 0$ and so $\Delta\xi_{t-n_o} = 0$, which implies that $\Delta\xi_t = 0$.

Consider the case $t > n_o$. Because

$$\|\Delta\xi_{t-n_o}\| \leq \|\xi_{t-n_o}(\alpha)\| + \|\xi_{t-n_o}(\beta)\|,$$

application of (6.6.2-8) to (6.6.2-10) leads to the inequality:

(6.6.2-11) $\|\Delta\xi_t\| \leq \prod_{\tau=1}^{n_o}\lambda(1+c_1\chi_1(t-\tau)+c_2\chi_2(t-\tau)).2c_3K'\lambda_1^{-n_o}$.

Applying (6.6.2-6) and (6.6.2-7) leads to

(6.6.2-12) $\|\Delta\xi_t\| \leq \min(\delta', \delta'')$.

Now $\Delta\xi_t$ will be related to $\Delta b_{t\omega} := b_{t\omega}(\alpha)-b_{t\omega}(\beta)$.

This can be done as follows; $\Delta b_{t\omega} = g_1g_2R^{-1}\Delta h_t$, so

(6.6.2-13) $\|\Delta b_{t\omega}\| \leq \bar{K}(R^{-1})\|\Delta h_t\|$.

From the definition of h it follows that

(6.6.2-14) $\|\Delta h_t\| \leq \sum_{i=1}^{d}\{|\Delta\psi_{it}^T\varepsilon_t| + |\psi_{it}^T\Delta\varepsilon_t| + |\Delta\psi_{it}^T\Delta\varepsilon_i|\}$.

From $z_t = H(\theta_t^+)\xi_t$ it follows that $\Delta z_t = H(\theta_t^+)\Delta\varepsilon_t$.

So making use of theorem (6.4-57),

(6.6.2-15) $\|z_t\| \leq \bar{K}(H)\|\xi_t\| \leq \bar{K}(H)c_3v_t \leq \bar{K}(H)c_3K'$,

and

(6.6.2-16) $\|\Delta z_t\| \leq \bar{K}(H)\|\Delta \xi_t\|$.

This implies

$$\|\Delta h_t\| \leq 2d\|\Delta z_t\|\|z_t\| + d\|\Delta z_t\|^2 \leq 2d\bar{K}(H)^2 c_3 K' \delta' + d\bar{K}(H)^2(\delta'')^2 \leq \tag{6.6.2-17}$$
$$\leq (\frac{\varepsilon'}{8} + \frac{\varepsilon'}{8})\bar{K}(R^{-1})^{-1}.$$

Combining this with (6.6.2-13) gives

(6.6.2-18) $\|\Delta b_{t\omega}\| < \varepsilon'/4$.

(ii) Fix all but the first n_o+1 components of the parameter sequence. I.e. Let $\{\theta^+_{t-\tau}\}_{\tau=n_o+1}^{\infty}$ be fixed. Then $b_{t\omega}$ can be considered as a function of $\{\theta^+_{t-\tau}\}_{\tau=o}^{n_o}$. The domain of this function is

(6.6.2-19) $\overset{\circ}{D}_b := \{(\theta^+_{t-\tau})_{\tau=o}^{n_o} | (\theta^+_{t-\tau})_{\tau=o}^{\infty} \in \tilde{\Pi}_t^a(\overset{\circ}{L}{}^+_{cP}(-\infty,0])\}$,

so $\overset{\circ}{D}_b \subseteq (\theta^+)^{n_o+1}$.

Note that $\overset{\circ}{D}_b$ is closed and therefore compact. The following property will be shown. For a given $\varepsilon' > 0$ a $\delta''' > 0$ can be found such that for all

$\alpha,\beta \in (\theta^+)^{\mathbb{N}_o}$ with $(\alpha_o,\alpha_1,\ldots,\alpha_{n_o}) \in \overset{\circ}{D}_b, (\beta_o,\beta_1,\ldots,\beta_{n_o}) \in \overset{\circ}{D}_b$ and

$\{\alpha_{n_o+i}\}_{i=1}^{\infty} = \{\beta_{n_o+i}\}_{i=1}^{\infty} = \{\theta^+_{t-n_o-i}\}_{i=1}^{\infty}$, the following implication holds

(6.6.2-20) $(\forall \tau \in \{0,1,\ldots,n_o\}: d^+(\alpha_\tau,\beta_\tau) < \delta''') \Rightarrow (\forall \omega: \|b_{t\omega}(\alpha) - b_{t\omega}(\beta)\| < \frac{1}{2}\varepsilon')$,

with d^+ as in (6.5-12). Note that δ''' may depend on the choice of the fixed part $\{\theta^+_{t-\tau}\}_{\tau=n_o+1}^{\infty}$ of the parameter sequence.

Because θ^+ is a disjoint union $\theta^+ = \bigcup_{\substack{i\in I(j)\\ j\in J}} \bar{D}'^+_{ij}$ of compact subsets $\bar{D}^+_{ij}$, the Cartesian product $(\theta^+)^{n_o+1}$ is a disjoint union of Cartesian products

$\bar{D}'^{+}_{i_o j_o} \times \bar{D}'^{+}_{i_1 j_1} \times \ldots \times \bar{D}'^{+}_{i_{n_o} j_{n_o}}$ with

$$\{(i_o,j_o),(i_1,j_1),\ldots,(i_{n_o},j_{n_o})\} \in \{(i,j) \mid i \in I(j), j \in J\}^{n_o+1},$$

and this last set is finite. Each set $\overset{\circ}{D}_b \cap (\bar{D}'^{+}_{i_o j_o} \times \ldots \times \bar{D}'^{+}_{i_{n_o} j_{n_o}})$ is compact and so continuity of a function on each such set implies uniform continuity. Let $((i_o,j_o),\ldots,(i_{n_o},j_{n_o}))$ be fixed. Just as in (i), if ω is such that $v_t > K'$ then $b_{t\omega}(\gamma) = 0$ for all γ, so $\|b_{t\omega}(\alpha)-b_{t\omega}(\beta)\| = 0 < \frac{1}{2}\varepsilon'$ in this case. Now consider the case $v_t \leq K'$. Then the inequality (6.6.2-8) holds. It implies

(6.6.2-21) $\forall \tau \in \{0,1,\ldots,n_o\}: \|\xi_{t-\tau}\| \leq c_3 K' \lambda_1^{-n_o}$.

So the vectors $\xi_t,\xi_{t-1},\ldots,\xi_{t-n_o}$ are uniformly bounded by $c_3 K' \lambda_1^{-n_o}$, independently of t,ω (provided $v_t \leq K'$) and independently of the choice of the parameter sequence. A similar argument can be set up for the observations y_t. Because $v_t \leq K'$ one has by definition of v_t, $\|y_t\| \leq v_t \leq K'$, and more generally

(6.6.2-22) $\|y_{t-\tau}\| \leq v_t \lambda_1^{-\tau} \leq K' \lambda_1^{-\tau}$, $\tau = 0,1,2,\ldots$

and so

(6.6.2-23) $\forall \tau \in \{0,1,2,\ldots n_o\}: \|y_{t-\tau}\| \leq K' \lambda_1^{-n_o}$,

independently of t,ω (provided $v_t \leq K'$). From this the property (6.6.2-20) can be proved in a straightforward manner, by writing out the formula for $b_{t\omega}$ in terms of $\theta^+_t,\theta^+_{t-1},\ldots,\theta^+_{t-n_o}$; $y_t,y_{t-1},\ldots,y_{t-n_o}$ and $(\xi_t,\xi_{t-1},\ldots,\xi_{t-n_o+1}),\xi_{t-n_o}$. Using the fact that $g_2(t) = 1$, $\forall t \geq t_3$ (so g_2 cannot cause discontinuities here) and using the upperbounds for the $\|\xi_{t-\tau}\|$ and the $\|y_{t-\tau}\|$, $\tau = 0,1,2,\ldots,n_o$ and using compactness of the parameter space,

(6.6.2-20) follows. The details are left to the reader.

(iii) Combining (i) and (ii) can be done as follows:
Suppose $\rho^+(\alpha,\beta) < \min((1+n_o)^{-1},\delta''') =: \delta$, with $\alpha,\beta \in \tilde{\Pi}^a_t(\mathring{L}^+_{cP}(-\infty,0])$.
Consider $\alpha = (\alpha^+_o,\alpha^+_1,\alpha^+_2,\ldots,\alpha^+_{n_o},\alpha^+_{n_o+1},\ldots)$,

$$\tilde{\alpha} := (\alpha^+_o,\alpha^+_1,\alpha^+_2,\ldots,\alpha^+_{n_o},\alpha^+_{n_o},\alpha^+_{n_o}\ldots),$$

$$\gamma := (\beta^+_o,\beta^+_1,\ldots,\beta^+_{n_o},\alpha^+_{n_o},\alpha^+_{n_o},\alpha^+_{n_o}\ldots),$$

and $\beta = (\beta^+_o,\beta^+_1,\ldots,\beta^+_{n_o},\beta^+_{n_o+1},\ldots)$.

Because $\alpha \in \tilde{\Pi}^a_t(\mathring{L}^+_{cP}(-\infty,0])$ it follows easily that $\tilde{\alpha} \in \tilde{\Pi}^a_t(\mathring{L}^+_{cP}(-\infty,0])$.

Now recall that n_o has been chosen, by definition, such that $n_o^{-1} \leq c\bar{a}_{t_n}$, so $\delta < n_o^{-1} \leq c\bar{a}_{t_4}$. It follows that $d^+(\alpha^+_{n_o},\beta^+_{n_o}) < c\bar{a}_{t_4}$ and that $\alpha^+_{n_o},\beta^+_{n_o}$ lie in the same coordinate chart $\bar{D}'^+_{ij}$.

Using also that $\alpha,\beta \in \tilde{\Pi}^a_t(\mathring{L}^+_{cP}(-\infty,0])$ it follows that $\gamma \in \tilde{\Pi}^a_t(\mathring{L}^+_{cP}(-\infty,0])$!

Therefore (i) and (ii) can be applied to the pairs $\alpha,\tilde{\alpha};\tilde{\alpha},\gamma$, and γ,β:

$$\|b_{t\omega}(\alpha)-b_{t\omega}(\beta)\| \leq \|b_{t\omega}(\alpha) - b_{t\omega}(\tilde{\alpha})\| + \|b_{t\omega}(\tilde{\alpha}) - b_{t\omega}(\gamma)\| +$$

$$\|b_{t\omega}(\gamma) - b_{t\omega}(\beta)\| < \frac{\varepsilon'}{4} + \frac{\varepsilon'}{2} + \frac{\varepsilon'}{4} = \varepsilon'. \qquad \text{Q.E.D.}$$

6.6.2-24. Theorem. Let t_3 be as in corollary (6.4-55).
The set of functions

$$\{b_{t\omega} \circ \Pi^a_t : \mathring{L}^+_{cP}(-\infty,0] \to \mathbb{R}^d \mid t \geq t_3, \omega \in \Omega\}$$

is equicontinuous.

Proof. This follows immediately from lemma (6.6.2-3) and proposition (6.5-14).
Q.E.D.

Let S_{-r}, $r \geq 0$ be the shift defined by

$$S_{-r}: \overset{\circ}{L}{}^{+}_{cP}(-\infty,0] \to \overset{\circ}{L}{}^{+}_{cP}(-\infty,0],$$

$$(S_{-r}Y)(s) = Y(s-r).$$

6.6.2-25. Corollary. Let $[b,c] \subseteq [0,\infty)$ be an arbitrary closed interval. The set of functions

$$\{b_{t\omega} \circ \Pi^{a}_{t} \circ S_{-r}: \overset{\circ}{L}{}^{+}_{cP}(-\infty,0] \to \mathbb{R}^{d} \mid t \geq t_3,\ r \in [b,c], \omega \in \Omega\}$$

is equicontinuous.

Proof. This follows from lemma (6.6.2-3) and proposition (6.5-14), using the fact that for all $s > 0$ and all $r \in (0,s)$ the following implication holds

$$\rho^{+}_{c}(Y^{+}_{1},Y^{+}_{2}) < \frac{1}{s} \Rightarrow \rho^{+}_{c}(Y^{+}_{1} \circ S_{-r}, Y^{+}_{2} \circ S_{-r}) < \frac{1}{s-r}.$$

The details are left to the reader.

Q.E.D.

Next the accumulated effects of the steps taken by the algorithm will be investigated. Consider sums of the form

$$(6.6.2\text{-}26) \qquad \sum_{\tau \in N_k} a_\tau b_\tau,$$

where the N_k are intervals such that the sums $\sum_{\tau \in N_k} a_\tau$ are approximately equal to a constant independent of k. Sums like (6.6.2-26) can and will be considered as integrals of piecewise constant functions, defined on the 'contracted time scale'. Let the partial sums of the series $\sum_{\tau=t_o}^{t-1} a_\tau$ be denoted by s_t, $t > t_o$ and let $s_{t_o} = 0$ and $s_t = -\sum_{\tau=t}^{t_o-1} a_\tau$ if $t < t_o$. Let the piecewise

constant function $b^o(r)$ be given by

$$(6.6.2\text{-}27)\quad \begin{cases} b^o(r) = b_{t_o} \text{ if } s_{t_o} = 0 \leq r < s_{t_o+1} = a_{t_o}, \\ b^o(r) = b_t \text{ if } s_t \leq r < s_{t+1} \text{ for some } t > t_o. \end{cases}$$

6.6.2-28. Proposition. Let $N = \{t \mid s_t \in [x,y]\}$, for some $x,y \in \mathbb{R}$, $x \leq y$. Let $t' = \min N$, $t'' = \max N$. Then

$$(6.6.2\text{-}29)\quad \int_x^y b^o(r)dr = \sum_{\tau \in N} a_\tau b_\tau + (s_{t'}-x)b_{t'-1} - (y-s_{t''+1})b_{t''}.$$

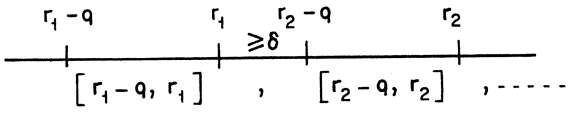
Proof.
$$\int_x^y b^o(r)dr = \int_x^{s_{t'}} b^o(r)dr + \int_{s_{t'}}^{s_{t''+1}} b^o(r)dr - \int_y^{s_{t''+1}} b^o(r)dr =$$

$$= (s_{t'}-x)b_{t'-1} + \sum_{\tau \in N} a_\tau b_\tau - (s_{t''+1}-y)b_{t''}.$$

Q.E.D.

To study the asymptotic behaviour use will be made of sequences of intervals, to be more precise: sequences of disjoint intervals in $[0,\infty)$, with monotonically increasing sequence of right endpoints, of equal interval length q and such that the distance between any two intervals is greater than or equal to δ for some predetermined $\delta > 0$. Each such sequence of intervals can be associated with a triple $\tilde{S} = (S,q,\delta)$ with $0 < \delta < q$ and $S = \{r_k\}_{k=1}^{\infty}$ a monotonically increasing sequence of positive numbers such that

$$(6.6.2\text{-}30)\quad \begin{cases} r_1 \geq q, \text{ and} \\ \forall k \geq 2\colon r_k - q \geq r_{k-1} + \delta. \end{cases}$$



6.6.2-31. Notation. The set of all such triples $\tilde{S}$ will be denoted by $\overset{\circ}{S}$.

6.6.2-32. Definition. Let $\{\theta_t^+ = (\theta_t, i_t, j_t)\}_{t=-\infty}^{T}$ be a parameter sequence with

$\theta_t^+ \in \Theta^+$ for each t, and $T \in \mathbb{Z}$ or $T = \infty$. Let $\tilde{S} \in \overset{\circ}{S}$. If $\ell \in \mathbb{N}$ is such that $s_{T+1} > r_\ell$, then the coordinate-chart-index set $I_\ell(\tilde{S},\{\theta_t^+\}_{t=-\infty}^{T})$ is defined by

(6.6.2-33) $$I_\ell(\tilde{S},\{\theta_t^+\}_{t=-\infty}^{T}) = \{(i_t,j_t) | s_t \in [r_\ell - q, r_\ell]\}.$$

If $T = \infty$ then the coordinate-chart-index set $I(\tilde{S},\{\theta_t^+\}_{-\infty}^{\infty})$ is defined as

(6.6.2-34) $$I(\tilde{S},\{\theta_t^+\}_{-\infty}^{\infty}) = \bigcup_{\ell\in\mathbb{N}} I_\ell(\tilde{S},\{\theta_t^+\}_{-\infty}^{\infty}).$$

Let $Y^+ \in \overset{\circ}{L}{}^+$, then $Y^+(r) = (Y(r),i(r),j(r))$. Let $\tilde{S} \in \overset{\circ}{S}$; $I_\ell(\tilde{S},Y^+)$ is defined by

$$I_\ell(\tilde{S},Y^+) = \{(i,j) | \exists r \in [r_\ell - q, r_\ell] \text{ such that } (i,j) = (i(r),j(r))\}$$

and

$$I(\tilde{S},Y^+) = \bigcup_{\ell=1}^{\infty} I_\ell(\tilde{S},Y^+).$$

6.6.2-35. Remarks. (i) If one deals with a parameter sequence $\{\theta_t^+\}_{t_o}^{T}$, $(T \geq t_o, T \in \mathbb{Z}$ or $T = \infty)$ one can formally extend it to $-\infty$ by taking $\theta_t^+ = \theta_{t_o}^+$ for all $t < t_o$, and apply the previous definition to the resulting sequence.
(ii) $I(\tilde{S},\{\theta_t^+\})$ and $I(\tilde{S},Y^+)$ can be interpreted as the set of all the 'relevant' coordinate-chart indices for the combination $(\tilde{S},\{\theta_t^+\})$ resp. $(\tilde{S},Y^+)$.

6.6.2-36. Definition. Let $\tilde{S} \in \overset{\circ}{S}$ be fixed. For each $k \in \mathbb{N}$, for each $\omega \in \Omega$ and for each $\{\theta_t^+\}_{t=-\infty}^{T}$ with $T \in \mathbb{Z}$ or $T = \infty$ such that $s_{T+1} > r_k$,

let $f_k = f_{k\omega}(\{\theta_t^+\}_{-\infty}^{T})$ be given by

(6.6.2-37) $$f_k = \begin{cases} 0 & \text{if } |I_k| > 1, \text{ and} \\ \int_{r_k - q + \delta}^{r_k} b^o(r)dr, & \text{if } |I_k| = 1, \end{cases}$$

where $I_k = I_k(\tilde{S},\{\theta_t^+\}_{-\infty}^{T})$.
For fixed $\tilde{S},\omega$ and k, f_k can be considered as a 'function' of the interpolation

curve $Y^+ \in L^+_{cP}(-\infty,0]$ (just like the variables $b_{t\omega}$). To state this precisely, let Π^r (or Π^r_a with $a = \{a_t\}$ if we want to stress the dependence on the sequence a) be defined by

$$\Pi^r : \mathring{L}^+_{cP}(-\infty,0] \to (\Theta^+)^{\mathbb{N}_o}$$

(6.6.2-38)
$$\Pi^r(Y^+) = \{\theta^+_{t-\tau}\}_{\tau=o}^{\infty} \text{ with } \theta_{t-\tau} = Y^+(s_{t-\tau}-r)$$

and t such that $0 \leq r-s_t < a_t$ (compare (6.5-9)). Analogously to (6.5-11) one has

6.6.2-39. Theorem. For each r and each $\tilde{a} = \{\tilde{a}_t\}$ with $\tilde{a}_t \geq 0$ for all t, $\Pi^r_{\tilde{a}}$ is continuous.

The proof is left to the reader.

Now $f_{k\omega} \circ \Pi^{r_k}$ is a function on $\mathring{L}^+_{cP}(-\infty,o]$ for each k and ω (for fixed $\tilde{S}$). The following result is a basic one of this section

6.6.2-40. Theorem. Let $\tilde{S} \in \mathring{S}$ be fixed. The set of functions

$$\{f_{k\omega} \circ \Pi^{r_k} : \mathring{L}^+_{cP}(-\infty,0] \to \mathbb{R}^d \mid k \in \mathbb{N},\ \omega \in \Omega\}$$

is equicontinuous.

Proof. For each $k \in \mathbb{N}$, $f_k \circ \Pi^{r_k}(Y^+)$ is a weighted sum of d-vectors $b_\ell \circ \Pi^{s_\ell}(Y^+) \circ \mathring{S}_{s_\ell - r_k}$ of which the (positive) weights are adding up to $q-\delta$ (= the length of the interval of integration). Therefore it is sufficient to show that these $b_\ell \circ \Pi^{s_\ell} \circ S_{s_\ell - r_k}$ are equicontinuous. Here $S_{s_\ell - r_k}$ is the shift over a length $r_k - s_\ell$. This length is bounded, because

$0 \leq r_k - s_\ell < q-\delta + \max_k a_k$ holds for all ℓ for which b_ℓ appears with a positive weight in the definition of f_k. Therefore corollary (6.6.2-25) applies and the result follows. Q.E.D.

It will be important for our purposes to show that asymptotically, for $k \to \infty$, the functions $f_k \circ \Pi^{r_k}(Y^+)$ depend only on the behaviour of the interpolation function Y^+ on the interval $[-q,0]$.

6.6.2-41. Theorem.
Let $\tilde{S} = (S,q,\delta) \in \overset{\circ}{S}$. Then $\forall \varepsilon > 0$, $\exists k_2 \in \mathbb{N}$ such that $\forall k \geq k_2, \forall \omega \in \Omega$ the following holds: if $Y_1^+, Y_2^+ \in \overset{\circ}{L}{}^+_{cP}(-\infty,0]$ and $Y_1^+|_{[q,0]} = Y_2^+|_{[-q,0]}$, then

$$\|f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\| < \varepsilon.$$

Proof. Let $\varepsilon > 0$ be given. From lemma (6.6.2-3) it follows that there exists an $n_o \in \mathbb{N}$ such that for all $t \geq t_3$ and for all $\omega \in \Omega$, $\forall \alpha, \beta \in \Pi_t(\overset{\circ}{L}{}^+_{cP}(-\infty,0])$ with $\alpha_i = \beta_i$, $i = 0,1,2,3,\ldots,n_o$, one has

(6.6.2-42) $\|b_{t\omega}(\alpha) - b_{t\omega}(\beta)\| < \varepsilon q^{-1}$.

Now let k_2 be such that for all $k \geq k_2$ the following three conditions hold simultaneously. Let $\tau_k := \min\{\tau \mid s_\tau \geq r_k - q + \delta\}$.

(i) $\tau_k > t_3$ (i.e. $r_k - q + \delta > s_{t_3}$).

(ii) $\operatorname{Sup}_{t \geq \tau_k - 1} a_t < \min(\frac{\delta}{2}, q - \delta)$

(this implies $r_k - q + \delta/2 \leq s_{\tau_k - 1} < r_k - q + \delta$ and $s_{\tau_k} < r_k$).

(iii) $s_{\tau_k - 1} - s_{\tau_k - 1 - n_o} < \frac{\delta}{2}$ (this implies, using (ii),

that $s_{\tau_k - 1 - n_o} > r_k - q$).

It follows that for all $k \geq k_2$ if $Y_1^+, Y_2^+ \in \overset{\circ}{L}{}^+_{cP}(-\infty,0]$ and $Y_1^+|_{[-q,0]} = Y_2^+|_{[-q,0]}$, then for all t with $s_t \in [r_k - q + \frac{\delta}{2}, r_k]$, the

sequences $\Pi_t \circ S_{-(r_k - s_t)}(Y_\ell^+)$ with $\ell = 1,2$ have equal first n_o terms:

$$\Pi_t \circ S_{-(r_k - s_t)}(Y_\ell^+) = \{Y_\ell^+(-\sum_{j=t-1}^{t-1} a_j + s_t - r_k)\}_{i=o}^{\infty} = \{Y_\ell^+(s_{t-i} - r_k)\}_{i=o}^{\infty},$$

according to (6.5-9) and the definition of s_t (just before (6.6.2-27)). Because $s_t \in [r_k-q+\frac{\delta}{2}, r_k]$, it follows from (i), (ii) and (iii) that for $i = 0,1,\ldots,n_o$, one has $s_{t-i}-r_k \in [-q,0]$, and so for all $i = 0,1,2,\ldots,n_o$, $Y_1^+(s_{t-i}-r_k) = Y_2^+(s_{t-i}-r_k)$. Combining this with (6.6.2-42) one obtains for all t with $s_t \in [r_k-q+\delta, r_k]$,

$$\text{(6.6.2-43)} \quad \|b_{t\omega} \circ \Pi_t \circ S_{-(r_k-s_t)}(Y_1^+) - b_{t\omega} \circ \Pi_t \circ S_{-(r_k-s_t)}(Y_2^+)\| < \varepsilon q^{-1}.$$

What does all this imply for the expression $\|f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\|$?

First notice that $I_k(\tilde{S}, \Pi^{r_k}(Y_1^+)) = I_k(\tilde{S}, \Pi^{r_k}(Y_2^+))$ because $Y_1^+|_{[-q,0]} = Y_2^+|_{[-q,0]}$. Therefore there are only two possibilities, namely (a) $|I_k| > 1$ in both cases (i.e. for Y_1^+ and Y_2^+) or (b) $|I_k| = 1$ in both cases. In case (a), according to (6.6.2-37),

$$\|f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\| = 0,$$

and in case (b), combining (6.6.2-37) with (6.6.2-43),

$$\|f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\| =$$

$$= \| \int_{r_k-q+\delta}^{r_k} (b^1(r)-b^2(r))dr \| < (q-\delta)\varepsilon q^{-1} < \varepsilon,$$

where $b^\ell(r)$, $\ell = 1,2$ is defined just as $b^o(r)$ in (6.6.2-27), with as parameter sequence $\Pi^{r_k}(Y_\ell^+)$, $\ell = 1,2$ respectively.

Q.E.D.

6.6.3. On the relation between the coupled and the decoupled algorithm. Having derived the necessary equicontinuity properties of the relevant variables in the previous subsections, we are now ready to apply the ideas that were introduced in section 6.6.1.

6.6.3-1. Definition. The probability measure P on $(\Omega,\mathcal{H})$ (i.e. on the data) induces a probability measure on the variables of the decoupled algorithm (described at the beginning of section 6.4) and all variables derived from those, like $\{f_k\}$ and $\{b_\ell\}$. This probability measure will be denoted by $\tilde{P}$. The

corresponding expectation operator will be denoted by $\tilde{E}$.

6.6.3-2. Remarks. (i) Because the decoupled algorithm is only well-defined if the parameter sequence $\{\theta_t^+ | \theta_t^+ \in \Theta^+\}$ is specified the same holds for $\tilde{P}$ and $\tilde{E}$. (ii) In the coupled algorithm, $\tilde{P}$ and $\tilde{E}$ are still well-defined. Because the result of taking expectations with respect to $\tilde{P}$ depends on the sequence of parameters and this sequence of parameters is now data-dependent, an expected value with respect to $\tilde{P}$ is also data-dependent (in general).

6.6.3-3. Theorem. Let $\tilde{S} \in \overset{\circ}{S}$ be fixed. The set of functions

$$\{\tilde{E}f_k \circ \Pi^{r_k} : \overset{\circ}{L}{}^+_{cP}(-\infty,0] \to \mathbb{R}^d \,|\, k \in \mathbb{N},\ \omega \in \Omega\}$$

is equicontinuous.

Proof. This follows directly from theorem (6.6.2-40). Indeed let $\varepsilon < 0$. There exists a $\delta > 0$ such that if $\rho_c^+(Y_1^+, Y_2^+) < \delta$, then for all k and for all $\omega \in \Omega$:

$$\|f_{k\omega} \circ \Pi^{r_k}(Y_1^+) - f_{k\omega} \circ \Pi^{r_k}(Y_2^+)\| < \varepsilon.$$

This implies that for all k and for all $\omega \in \Omega$:

$$\|\tilde{E}f_k \circ \Pi^{r_k}(Y_1^+) - \tilde{E}f_k \circ \Pi^{r_k}(Y_2^+)\| =$$

$$= \|\tilde{E}\{f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\}\| \leq$$

$$\leq \tilde{E}\|f_k \circ \Pi^{r_k}(Y_1^+) - f_k \circ \Pi^{r_k}(Y_2^+)\| \leq \tilde{E}\varepsilon = \varepsilon.$$

Q.E.D.

6.6.3-4. Corollary. Let $\tilde{S} \in \overset{\circ}{S}$ be fixed. The set of functions

$$\{f_{k\omega} - \tilde{E}f_k : \overset{\circ}{L}{}^+_{cP}(-\infty,0] \to \mathbb{R}^d \,|\, k \in \mathbb{N}, \omega \in \Omega\}$$

is equicontinuous.

The proof follows immediately from theorems (6.6.2-40) and (6.6.3-3).

In this section the result will be derived that if for each parameter sequence $\{\theta_t^+\}$ satisfying property 3 (cf. (6.5-17)) - with interpolation function

$Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta^+_t = Y^+(s_t)$ for all $t \in \mathbb{Z}$ - and for each $\tilde{S}$ for which

$|I(\tilde{S},Y^+)| = 1$ (and hence $|I(\tilde{S},\{\theta^+_t\})| = 1$), one has

$$\lim_{k\to\infty} f_k - \tilde{E}f_k = 0, \ \tilde{P}\text{-a.s.},$$

then for the coupled algorithm one has

$$(\forall \tilde{S} \in \overset{\circ}{S} \text{ with } |I(\tilde{S},Y^+)| = 1, \lim_{k\to\infty} f_k - \tilde{E}f_k = 0), \ P\text{-a.s.}$$

First let us formulate precisely the hypothesis concerning $\tilde{P}$-a.s. convergence of $f_k - \tilde{E}f_k$. (The hypothesis will be proven to be true in section 6.7.2. Here we are only dealing with its implications for the coupled algorithm).

6.6.3-5. Hypothesis. For all parameter sequences $\{\theta^+_t\}$ satisfying property 3 - with interpolation function $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta^+_t = Y^+(s_t), \forall t \in \mathbb{Z}$ - and for all $\tilde{S} \in \overset{\circ}{S}$ such that $|I(\tilde{S},Y^+)| = 1$, there exists a subset N of Ω with $P(N) = 0$, such that for all $\omega \in \Omega\backslash N$,

$$\lim_{k\to\infty} f_{k\omega} - \tilde{E}f_k = 0.$$

6.6.3-6. Theorem. Let $\tilde{S} \in \overset{\circ}{S}$ be fixed. Suppose the hypothesis (6.6.3-5) holds. Then there exists a subset N of Ω with $P(N) = 0$ such that for each parameter sequence $\{\theta^+_t\}$ satisfying property 3 - with interpolation function $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta^+_t = Y^+(s_t)$ for all $t \in \mathbb{Z}$ - and for each subsequence $\bar{S} = \{r_{k(j)}\}_{j=1}^{\infty}$ of $S = \{r_k\}_{k=1}^{\infty}$ for which

$$|I((\bar{S},q,\delta), Y^+)| = 1,$$

one has

$$\forall \omega \in \Omega\backslash N: \lim_{j\to\infty} f_{k(j)} - \tilde{E}f_{k(j)} = 0.$$

Proof. The proof is rather long. Let $\{\bar{f}_k\}_{k=1}^{\infty}$ be defined by

$$(6.6.3\text{-}7) \qquad \bar{f}_k = \begin{cases} 0 & \text{if } \forall j \in \mathbb{N}: k(j) \neq k, \\ f_{k(j)} & \text{if } \exists j \in \mathbb{N}: k(j) = k. \end{cases}$$

It is clear that

$$\lim_{k\to\infty} \bar{f}_k - \tilde{\bar{E}}f_k = 0 \text{ iff } \lim_{j\to\infty} f_{k(j)} - \tilde{E}f_{k(j)} = 0.$$

Because $|I((\bar{S},q,\delta),Y^+)| = 1$ there is only a finite number of possibilities for $I((\bar{S},q,\delta),Y^+)$. Therefore it is sufficient to prove the result for a fixed index $(i,j) = I((\bar{S},q,\delta),Y^+)$, because a finite union of null sets (i.e. sets of measure zero) is a null set.

In section 6.5 we saw that the spaces $\overset{\circ}{L}_{ij}[a,b]$, with $[a,b] \subseteq \mathbb{R}$, are compact (theorem (6.5-7)(b)). The idea of the proof is now to relate the functions $\bar{f}_k \circ \Pi^{r_k}: \overset{\circ}{L}{}^+_{cP} \to \mathbb{R}^d$ to functions $f_k', f_k'': \overset{\circ}{L}{}^+_{ij}[-q,0] \to \mathbb{R}^d$ using theorem (6.6.2-41), and then to use theorem (6.6.1-8).

The functions f_k', $k \in \mathbb{N}$, are defined as follows. To each $Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ associate $Y' \in \overset{\circ}{L}{}^+_{ij}(-\infty,0]$, defined by

$$(6.6.3\text{-}8) \quad \begin{cases} Y'|_{[-q,0]} = Y|_{[-q,0]} \text{ and} \\ Y'(s) = Y(-q) \text{ for all } s < -q. \end{cases}$$

Note that indeed $Y' \in \overset{\circ}{L}{}^+_{ij}(-\infty,0]$; i.e. it does satisfy the Lipschitz condition and it remains in the coordinate chart $\bar{D}_{ij}'$. So $Y' \in \overset{\circ}{L}{}^+_{ij}(-\infty,0] \subseteq \overset{\circ}{L}{}^+_{cP}(-\infty,0]$.

Now let

$$(6.6.3\text{-}9) \quad f_k'(Y) := f_k \circ \Pi^{r_k}(Y'), \ \forall k \in \mathbb{N}.$$

The functions f_k'', $k \in \mathbb{N}$, are defined as follows. With each element $Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ associate an element $Y'' \in \overset{\circ}{L}{}^+_{ij}$ as follows.

Let Y^{-1} denote the 'reverse' of Y

$$Y^{-1}(t) := Y(-q-t), \quad \forall t \in [-q,0].$$

Let

$$(6.6.3\text{-}10)\quad \begin{cases} Y''(s)\big|_{[r_k-q,\,r_k]} = \begin{cases} Y(s-r_k) \text{ if } k \text{ is odd,} \\ Y^{-1}(s-r_k) \text{ if } k \text{ is even,} \end{cases} \\ Y''\big|_{[r_k,\,r_{k+1}-q]} = \text{constant} = \begin{cases} Y(q) \text{ if } k \text{ is odd,} \\ Y(0) \text{ if } k \text{ is even and} \end{cases} \\ Y''\big|_{(-\infty,\,r_1-q]} = Y(0). \end{cases}$$

Note that indeed $Y'' \in \overset{\circ}{L}{}^{+}_{ij}$ (the Lipschitz condition holds and $Y''(s) \in \bar{D}'^{+}_{ij}$ for all $s \in \mathbb{R}$). Let

$$(6.6.3\text{-}11)\quad f_k''(Y) := f_k(\{Y''(s_t)\}_{t=-\infty}^{\infty}) = f_k \circ \Pi^{r_k} \circ (S_{-r_k}(Y''))\big|_{(-\infty,0]},\ \forall k \in \mathbb{N}.$$

Consider the following four assertions

(i) $\forall Y \in \overset{\circ}{L}{}^{+}_{ij}[-q,0]$: $\exists N, P(N) = 0$ such that

$$\forall \omega \in \Omega\backslash N\text{: } \lim_{k\to\infty} f_k''(Y) - \tilde{E}f_k''(Y) = 0,$$

(ii) $\forall Y \in \overset{\circ}{L}{}^{+}_{ij}[-q,0]$: $\exists N$, $P(N) = 0$ such that

$$\forall \omega \in \Omega\backslash N\text{: } \lim_{k\to\infty} f_k'(Y) - \tilde{E}f_k'(Y) = 0,$$

(iii) $\exists N, P(N) = 0$ such that $\forall \{Y_k \,|\, Y_k \in \overset{\circ}{L}{}^{+}_{ij}[-q,0]\}_{k=1}^{\infty}$, $\forall \omega \in \Omega\backslash N$:

$$\lim_{k\to\infty} f_k'(Y_k) - \tilde{E}f_k'(Y_k) = 0,$$

(iv) $\exists N, P(N) = 0$ such that for all $\omega \in \Omega\backslash N$, for all $\{\theta_t^+\}$ satisfying property 3 and for all subsequences $\bar{S}$ of $\overset{\circ}{S}$ such that $I((\bar{S},q,\delta),\{\theta_t^+\}) = \{(i,j)\}$:

$$\lim_{k\to\infty} \bar{f}_k - \tilde{\bar{E}}\bar{f}_k = 0.$$

The following sequence of implications will be shown (making use of theorems derived before):

Hypothesis $\overset{\Rightarrow}{(a)}$(i) $\overset{\Rightarrow}{(b)}$(ii) $\overset{\Rightarrow}{(c)}$(iii) $\overset{\Rightarrow}{(d)}$(iv)

(a) For arbitrary $Y \in \overset{\circ}{L}{}^{+}_{ij}[-q,0]$, consider the parameter sequence $\{\theta_t^+ = Y''(s_t) \,|\, t \in \mathbb{Z}\}$. From (6.6.3-11) it is clear that

$$f_k''(Y) - \tilde{E}f_k''(Y) = f_k(\{\theta_t^+\}) - \tilde{E}f_k(\{\theta_t^+\}).$$

The hypothesis implies that there exists a set $N \subseteq \Omega$ with $P(N) = 0$, such that

$$\forall \omega \in \Omega\backslash N: \lim_{k\to\infty} f_k(\{\theta_t^+\}) - \tilde{E}f_k(\{\theta_t^+\}) = 0.$$

(Note that the coordinate chart index (i,j) is constant and equal to (i,j) for all t). So $\forall Y \in \overset{\circ}{L}{}^+_{ij}[-q,0], \exists N, P(N) = 0$ such that $\forall \omega \in \Omega\backslash N$: $f_k''(Y) - \tilde{E}f_k''(Y) = 0$, so the hypothesis implies (i) indeed.

(b) Let $Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ be arbitrary.
Because

$$\forall s \in [-q,0] \begin{cases} Y(s) = Y'(s) = Y''(s+r_k) \text{ if } k \text{ is even, and} \\ Y(s) = Y'(s) = Y''(r_k - q - s) = \\ \qquad = (Y^{-1})''(s+r_k) \text{ if } k \text{ is odd.} \end{cases}$$

Application of theorem (6.6.2-41) gives us

$$\forall \omega \in \Omega, \forall Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]:$$

(6.6.3-12)
$$\lim_{k\to\infty} f'_{2k}(Y) - f''_{2k}(Y) = 0 \text{ and } \lim_{k\to\infty} f'_{2k-1}(Y) - f''_{2k-1}(Y^{-1}) = 0,$$

and

$$\forall Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]:$$

(6.6.3-13)
$$\lim_{k\to\infty} \tilde{E}f'_{2k}(Y) - \tilde{E}f''_{2k}(Y) = 0 \text{ and } \lim_{k\to\infty} \tilde{E}f'_{2k-1}(Y) - \tilde{E}f''_{2k-1}(Y^{-1}) = 0.$$

So (i) implies (taking for each $Y \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ the union of the null sets for Y and Y^{-1}, which results in another null set) $\forall Y \in \overset{\circ}{L}{}^+_{ij}[-q,0], \exists N, P(N) = 0$ such that

$$\forall \omega \in \Omega\backslash N: \lim_{k\to\infty} f_k'(Y) - \tilde{E}f_k'(Y) = 0.$$

So indeed (i) implies (ii).

(c) Corollary (6.6.3-4) states that the set of functions

$$\{f_k - \tilde{E}f_k : \overset{\circ}{L}{}^{+}_{cP}(-\infty,0] \to \mathbb{R}^d \mid k \in \mathbb{N},\ \omega \in \Omega\}$$

is equicontinuous. The same holds a fortiori if the domain of the functions is restricted to a subset of $\overset{\circ}{L}{}^{+}_{cP}(-\infty,0]$. Let us restrict the domain to the set of all Y' with $Y \in \overset{\circ}{L}{}^{+}_{ij}[-q,0]$. From the equicontinuity of the resulting set of functions it follows that the set of functions

$$\{f_k' - \tilde{E}f_k' : \overset{\circ}{L}{}^{+}_{ij}[-q,0] \to \mathbb{R}^d \mid k \in \mathbb{N}, \omega \in \Omega\}$$

is equicontinuous as well. The space $\overset{\circ}{L}{}^{+}_{ij}[-q,0]$ is compact (cf. theorem (6.5-7)(b)). Therefore theorem (6.6.1-8)(a) is applicable; it tells us that (ii) implies (iii).

(d) Let $\{\theta_t^+\}$ be a sequence satisfying property 3, and let $\overset{\circ}{Y}{}^{+} \in \overset{\circ}{L}{}^{+}_{cP}$ be an interpolation curve, i.e. $\theta_t^+ = Y^+(s_t)$, $\forall t \in \mathbb{Z}$. Let $\bar{S} = \{r_{k(j)}\}_{j=1}^{\infty}$ be a subsequence of S such that $I((\bar{S},q,\delta),Y^+) = \{(i,j)\}$. Let for each $k \in \mathbb{N}$, $Y \in \overset{\circ}{L}_{cP}(-\infty,0]$ be defined by

$$Y_k(r) = Y^+(r_k + r),\ \forall r \in (-\infty,0].$$

Then $Y_{k(j)}|_{[-q,0]} \in \overset{\circ}{L}_{ij}[-q,0]$ for each $j \in \mathbb{N}$.

By construction of the Y_k, one has

$$\bar{f}_k(\{\theta_t^+\}) - \tilde{\bar{E}f}_k(\{\theta_t^+\}) = \bar{f}_k \circ \Pi^{r_k}(Y_k) - \tilde{\bar{E}f}_k \circ \Pi^{r_k}(Y_k),\ \forall k \in \mathbb{N}. \tag{6.6.3-14}$$

Now (iii), together with theorem (6.6.2-41) implies

$$\exists N, P(N) = 0, \text{ such that } \forall \omega \in \Omega \backslash N: \tag{6.6.3-15}$$

$$\lim_{k\to\infty} \bar{f}_k - \tilde{\bar{E}f}_k = \lim_{k\to\infty} \bar{f}_k \circ \Pi^{r_k}(Y_k) - \tilde{\bar{E}f} \circ \Pi^{r_k}(Y_k) =$$

$$= \lim_{k\to\infty} \bar{f}_k'(Y_k|_{[-q,0]}) - \tilde{\bar{E}f}_k'(Y_k|_{[-q,0]}) = 0.$$

Q.E.D.

The problem that is left is that the exceptions sets in the previous theorem depend on the choice of the sequence $\tilde{S} \in \overset{\circ}{S}$. And as the set $\overset{\circ}{S}$ is uncountable, one can not conclude directly that there exists one exceptions set N with $P(N) = 0$, which contains all the exceptions for all possible choices of $\tilde{S}$. It will, however, be shown that such a set exists.

6.6.3-16. Theorem. The hypothesis (6.6.3-5) implies: $\exists E \subseteq \Omega$ with $P(E) = 0$, such that for all $\{\theta_t^+\}$ satisfying property 3 - with an interpolation curve $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta_t^+ = Y^+(s_t), \forall t \in \mathbb{Z}$ - and for all $\tilde{S} \in \overset{\circ}{S}$ such that

$|I(\tilde{S},Y^+)| = 1$, one has

$$\forall \omega \in \Omega \backslash E: \lim_{k\to\infty} f_k - \tilde{E}f_k = 0.$$

Before giving the proof let us state a corollary.

6.6.3-17. Corollary. The hypothesis (6.6.3-5) implies the following. Suppose $\{\theta_t^+(\omega)\}$ is an ω-dependent parameter sequence, satisfying property 3 and with interpolation curve Y_ω^+ as in the previous theorem. Then

$$\exists E, P(E) = 0, \forall \omega \notin E, \forall \tilde{S} \in S \text{ with } |I(\tilde{S},Y_\omega^+)| = 1, \lim_{k\to\infty} f_k - \tilde{E}f_k = 0.$$

Proof of the theorem.

Let $\bar{P} = \{(p_1,p_2) | p_1 \in Q, p_2 \in Q, p_1 > p_2 > 0\}$. Then $\bar{P}$ is countable. For each $p \in \bar{P}$, let us define

$$(6.6.3\text{-}18) \quad \tilde{S}_p = (S_p,p_1,p_2) := (\{(p_1+p_2)\ell\}_{\ell=1}^{\infty},p_1,p_2).$$

Then $\{\tilde{S}_p | p \in \bar{P}\}$ forms a countable subset of $\overset{\circ}{S}$. Application of theorem (6.6.3-6) to $\tilde{S}_p$ leads to an exceptions set that will be denoted by E_p (in the theorem it is denoted by N), with the property $P(E_p) = 0$. Now let

$$(6.6.3\text{-}19) \quad E = \bigcup_{p\in\bar{P}} E_p.$$

Then $P(E) = 0$ because $\bar{P}$ is countable. Note that the intervals corresponding to $\tilde{S}_p$ are

(6.6.3-20) $[(p_1+p_2)(\ell-1) + 2p_2,(p_1+p_2)\ell]$, $\ell \in \mathbb{N}$.

Let $\{\theta_t^+\}$ be a parameter sequence satisfying property 3 and $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ a corresponding interpolation function as before. Let $\tilde{S} = (S,q,\delta) = (\{r_k\},q,\delta)$ be some element of $\overset{\circ}{S}$ such that $|I(\tilde{S},Y^+)| = 1$. Let $\varepsilon > 0$. It will clearly be sufficient to show that

(6.6.3-21) $\forall \omega \in \Omega\setminus E$: $\limsup_{k\to\infty} \|f_k - \tilde{E}f_k\| < \varepsilon$.

Fix $p \in P$, in order to derive an inequality.
For each $k \in \mathbb{N}$, let

(6.6.3-22) $L_k := \{\ell\in\mathbb{N} \mid [(p_1+p_2)(\ell-1)+2p_2,(p_1+p_2)\ell] \subseteq [r_k-q+\delta,r_k]\}$;

notice that

(6.6.3-23) $(q-\delta-2p_1)/(p_1+p_2) \leq |L_k| \leq (q-\delta+2p_2)/(p_1+p_2)$.

Let for all $\ell \in L_k$

(6.6.3-24) $H_{k\ell} := [(p_1+p_2)(\ell-1)+2p_2,(p_1+p_2)\ell]$,

and

(6.6.3-25) $H_k := \bigcup_{\ell\in L_k} H_{k\ell}$.

Let λ denote Lebesgue measure. One has

(6.6.3-26) $\lambda([r_k-q+\delta,r_k]\setminus H_k) \leq (q-\delta) - |L_k|(p_1+p_2) \leq$

$$\leq (q-\delta) - \left(\frac{q-\delta-2p_1}{p_1+p_2}\right)(p_1-p_2) = \frac{(q-\delta)2\left(\frac{p_2}{p_1}\right)+p_1\left(1-\left(\frac{p_2}{p_1}\right)\right)}{1+\left(\frac{p_2}{p_1}\right)} \leq$$

$$\leq (q-\delta)2\left(\frac{p_2}{p_1}\right) + p_1.$$

Note that this is small if $\frac{p_2}{p_1}$ and p_1 are small enough.

Let $\bar{S}_p = \{(p_1+p_2)\ell(i)\}_{i=1}^{\infty}$ be the subsequence of S_p with the property

(6.6.3-27) $\{\ell(i) \mid i \in \mathbb{N}\} = \bigcup_{k=1}^{\infty} L_k$.

Let $\{f_i^p\}_{i=1}^{\infty}$ denote the corresponding f-sequence, i.e.

(6.6.3-28) $f_i^p = \int_{H_{k,\ell(i)}} b^o(\sigma)d\sigma$ with k such that $\ell(i) \in L_k$.

For each $k \in \mathbb{N}$, one has

$$\begin{aligned} \|f_k-\tilde{E}f_k\| &= \Big\| \int_{r_k-q+\delta}^{r_k} [b^o(\sigma)-\tilde{E}b^o(\sigma)]d\sigma \Big\| \leq \\ &\leq \Big\| \int_{[r_k-q+\delta,r_k]\setminus H_k} [b^o(\sigma)-\tilde{E}b^o(\sigma)]d\sigma \Big\| + \sum_{\ell\in L_k} \Big\| \int_{H_{k\ell}} [b^o(\sigma)-\tilde{E}b^o(\sigma)]d\sigma \Big\|. \end{aligned} \tag{6.6.3-29}$$

Let $i_k = \max\{i \mid \ell(i)\in L_k\}$ for each k. From section 6.4 (see e.g. the proof of corollary (6.4-55)) it follows that $b^o(\sigma)$ is bounded by a data-independent constant, because the parameter sequence satisfies property 3 and so a fortiori property 1. Let this constant be denoted by K_b. Then it follows that

(6.6.3-30) $\|f_k-\tilde{E}f_k\| \leq 2K_b.\lambda([r_k-q+\delta,r_k]\setminus H_k) + \sum_{i=i_k-|L_k|+1}^{i_k} \|f_i^p-\tilde{E}f_i^p\|$.

Applying theorem (6.6.3-6) to S_p, and using $E_p \subseteq E$, $P(E) = 0$, we find that the hypothesis (6.6.3-5) implies:

$$\forall \omega \in \Omega\setminus E: \lim_{i\to\infty} \|f_i^p-\tilde{E}f_i^p\| = 0.$$

Because $\{|L_k|\}_{k=1}^{\infty}$ is bounded (cf. (6.6.3-23)), and applying (6.6.3-26), one finds

(6.6.3-31) $\forall \omega \in \Omega\setminus E: \limsup_{k\to\infty} \|f_k-\tilde{E}f_k\| \leq 2K_b\{(q-\delta)2\frac{p_2}{p_1}+p_1\}$.

This is the inequality mentioned right after (6.6.3-21). It is now trivial to

see that $p = (p_1,p_2) \in \bar{P}$ can be chosen such that the right-hand side of (6.6.3-31) is smaller than $\varepsilon > 0$.

Q.E.D.

6.7. Exponential decay properties of the algorithm and the implications for convergence

6.7.1. About some exponential decay properties

The main purpose of this section 6.7 is to prove that the hypothesis (6.6.3-5) holds. To do this, our main tools will be the 'exponential decay properties' of the algorithm. In this first subsection definitions of several decay properties will be given and some results will be derived. Applications to the algorithm will be treated in section 6.7.2.

Let us start with defining exponential decay. In the following let B denote a d-dimensional vector space ($d \in \mathbb{N}$) with inner product $\langle,\rangle$ and norm $\|,\|$ and let the vectors of B be represented with respect to some orthonormal basis.

6.7.1-1. Definition. (i) A sequence $\{b_k | b_k \in B\}_{k=1}^{\infty}$ converges exponentially to zero or, equivalently, decays exponentially if the following holds

(6.7.1-2) $\quad \exists c > 0, \exists \lambda \in (0,1): \forall k \in \mathbb{N}: \|b_k\| < c\lambda^k.$

(ii) Let $\bar{N} = \{(k,\ell) | k \in \mathbb{N}, \ell \in \mathbb{N}, k \geq \ell\}$. A double sequence $\{b_{k\ell} | b_{k\ell} \in B\}_{(k,\ell)\in\bar{N}}$ is said to be exponentially decaying if

(6.7.1-3) $\quad \exists c > 0, \exists \lambda \in (0,1): \forall (k,\ell) \in \bar{N}, \|b_{k\ell}\| < c\lambda^{k-\ell}.$

Remark. In this definition $\lambda \in (0,1)$ can be replaced by $e^{-\alpha}$ with $\alpha > 0$. This explains the word 'exponentially'.

6.7.1-4. Notation. If $\{b_k\}_1^{\infty}$ satisfies (6.7.1-2) (i) we will say that '$\{b_k\}_1^{\infty}$ has e.d.' (exponential decay) or that '$\{b_k\}_1^{\infty}$ is an e.d. sequence' (exponentially decaying sequence). Similarly, if $\{b_{k\ell} | (k,\ell) \in \bar{N}\}$ satisfies (6.7.1-2) (ii) we will say that '$\{b_{k\ell} | (k,\ell) \in \bar{N}\}$ has e.d.' or that '$\{b_{k\ell} | (k,\ell) \in \bar{N}\}$ is an e.d. double sequence'.

6.7.1-5. Theorem. (i) Let $\{b_k\}_{k=1}^{\infty}$, $\{c_k\}_{k=1}^{\infty}$, $\{d_k^{(j)}\}_{k=1}^{\infty}$, $j \in \mathbb{N}$, be e.d. sequences. Then

(a) $\{b_k+c_k\}_1^\infty$ is an e.d. sequence,

(b) $\{b_k^T c_k\}_1^\infty$ is an e.d. sequence,

(c) $\{kb_k\}_1^\infty$ is an e.d. sequence,

(d) $\forall \lambda_1 \in (0,1)$: $\{b_k' = \sum_{j=o}^{k} \lambda_1^j b_{k-j}\}_{k=1}^\infty$ is an e.d. sequence,

(e) if the b_k are nonnegative scalars and $\mu > 0$ is arbitrary, then $\{b_k^\mu\}_{k=1}^\infty$ is an e.d. sequence, and

(f) if $p(x^{(1)},\ldots,x^{(n)})$ is a polynomial in n variables without constant term, i.e. $p(0,\ldots,0) = 0$, then the sequence $\{p(d_k^{(1)},\ldots,d_k^{(n)})\}_{k=1}^\infty$ has e.d.

(ii) Let $\{b_{k\ell}|(k,\ell)\in\bar{N}\}$, $\{c_{k\ell}|(k,\ell)\in\bar{N}\}$, $\{d_{k\ell}^{(j)}|(k,\ell)\in\bar{N}\}$, $j \in \mathbb{N}$, be e.d. double sequences. Then

(a) $\{b_{k\ell}+c_{k\ell}|(k,\ell)\in\bar{N}\}$ is an e.d. double sequence,

(b) $\{b_{k\ell}^T c_{k\ell}|(k,\ell) \in \bar{N}\}$ is an e.d. double sequence,

(c) $\{(k-\ell)b_{k\ell}|(k,\ell) \in \bar{N}\}$ is an e.d. double sequence,

(d) $\forall \lambda_1 \in (0,1)$: $\{b_{k\ell}' = \sum_{j=1}^{\ell} \lambda_1^{j-1} b_{kj}|(k,\ell) \in \bar{N}\}$ is an e.d. double sequence,

(e) if the $b_{k\ell}$ are nonnegative scalars and $\mu > 0$ is arbitrary, then $\{b_{k\ell}^\mu|(k,\ell) \in \bar{N}\}$ is an e.d. sequence, and

(f) if $p(x^{(1)},\ldots,x^{(n)})$ is a polynomial in n variables without constant term i.e. $p(0,\ldots,0) = 0$, then the double sequence $\{p(d_{k\ell}^{(1)},d_{k\ell}^{(2)},\ldots,d_{k\ell}^{(n)})|(k,\ell) \in \bar{N}\}$ has e.d.

Proof. We will give the proof of (i); the proof of (ii) is completely analogous.

(a) is trivial

(b) is trivial in the scalar case; the vector case follows easily by applying repeatedly (a) and the scalar case of (b).

(c) $\{b_k\}$ has e.d., so $\exists c > 0$, $\exists \lambda \in (0,1)$ such that $\forall k \in \mathbb{N}$: $\|b_k\| < c\lambda^k$ which implies $\forall k \in \mathbb{N}$: $\|kb_k\| < ck\lambda^k$.

Take $\lambda_2 \in (\lambda,1)$. Then $\lim_{k\to\infty} \frac{k\lambda^k}{\lambda_2^k} = 0$, so $\exists k_1 \in \mathbb{N}$ such that $\forall k \geq k_1$: $k\lambda^k \leq \lambda_2^k$. Now let

$$(6.7.1\text{-}6) \qquad c_2 := \max\Big(c, \max_{1\leq k\leq k_1} c.k.\frac{\lambda^k}{\lambda_1^k}\Big),$$

then

$$(6.7.1\text{-}7) \qquad \forall k \in \mathbb{N}: \|kb_k\| < c_2\lambda_2^k.$$

(d) Because $\{b_k\}$ has e.d., $\exists c_3 > 0$, $\exists \lambda_3 \in (0,1)$ such that $\|b_k\| < c_3\lambda_3^k$ for all $k \in \mathbb{N}$. Consider

$$\|b_k'\| = \Big\| \sum_{j=o}^{k-1} \lambda_1^j b_{k-j} \Big\| \leq \sum_{j=o}^{k-1} \lambda_1^j \lambda_3^{k-j} c_3 \leq k.\max(\lambda_1,\lambda_3)^k c_3,$$

and $\{k \max(\lambda_1,\lambda_3)^k c_3\}_{k=1}^{\infty}$ is an e.d. sequence according to (b). So $\{b_k'\}_1^{\infty}$ is an e.d. sequence.

(e) In this case one has

$$\exists c > 0, \ \exists \lambda \in (0,1), \ \forall k \in \mathbb{N}: \ 0 < b_k < c\lambda^k.$$

It follows that for arbitrary $\mu > 0$

$\forall k \in \mathbb{N}: 0 < b_k^\mu < c^\mu(\lambda^\mu)^k$, and $c^\mu > 0$, $\lambda^\mu \in (0,1)$ of course, so $\{b_k^\mu\}$ is an e.d. sequence.

(f) This follows easily by repeated application of (a) and the scalar case of (b). Q.E.D.

Let us now consider e.d. sequences of matrices. First let us show that for the definition it does not matter whether the spectral norm or the Frobenius norm is used.

6.7.1-8. Lemma. $\{A_k\}_{k=1}^{\infty}$ is an e.d. sequence of $d_1 \times d_2$ matrices with respect to the spectral norm iff it is an e.d. sequence with respect to the Frobenius norm.

The proof follows easily from the fact that for any matrix A,

(6.7.1-9) $\|A\|_S \leq \|A\|_F \leq \sqrt{n}.\|A\|_S$,

where $\|,\|_S$ is the spectral norm and $\|,\|_F$ the Frobenius norm, as before. By convention, with an e.d. sequence of matrices $\{A_k\}_1^\infty$ will be meant an e.d. sequence with respect to the Frobenius norm or equivalently with respect to the spectral norm (i.e. $\{\|A_k\|_F\}$ or equivalently $\{\|A_k\|_S\}$ is an e.d. sequence of numbers).

Remark. In the following the real case will be considered. The results for the complex case are analogous.

6.7.1-10. Lemma. $\{A_k\}_{k=1}^\infty$ is an e.d. sequence of $d_1 \times d_2$ (real) matrices iff for each d_2-vector, $\{A_k x\}_1^\infty$ is an e.d. sequence of d_1-vectors.

Proof. If $\{A_k\}_{k=1}^\infty$ is an e.d. sequence then $\exists c > 0, \exists \lambda \in (0,1)$: $\|A_k\|_S < c\lambda^k$, and so if $c_1 \geq c\|x\| > 0$, then $\|A_k x\| \leq \|A_k\|_S \|x\| < c_1 \lambda^k$, which shows that $\{A_k x\}$ has e.d. On the other hand, if $\{A_k x\}_{k=1}^\infty$ has e.d. for each d_2-vector x, then it follows easily that each component-sequence $\{e_i^T A_k e_j\}_{k=1}^\infty$, where e_i denotes the i^{th} unit vector, has e.d. and therefore $\{A_k\}$ is an e.d. sequence. The details are left to the reader.

Q.E.D.

6.7.1-11. Lemma. Let A be a square $d \times d$ matrix. $\{A^k\}_{k=1}^\infty$ is an e.d. sequence iff

A is asymptotically stable.

Proof. If $\{A^k\}$ is an e.d. sequence then clearly $\lim_{k\to\infty} A^k = 0$, and so all eigenvalues have modulus less than one, so A is asymptotically stable.
On the other hand if A is asymptotically stable then its eigenvalues lie in the open unit disk. Let $\lambda \in (0,1)$ be larger than the largest modulus of any eigenvalue of A. Then

(6.7.1-12) $$\lim_{k\to\infty} \|A^k\|_S / \lambda^k = 0,$$

and so $\{\|A^k\|_S / \lambda^k\}_{k=1}^\infty$ is bounded by some positive number c (say).

So $\forall k \in \mathbb{N}$: $\|A^k\|_S < c\lambda^k$, and so $\{A^k\}$ is e.d. Q.E.D.

Remark. Notice that '$\{A^k\}$ is an e.d. sequence' is definitely a weaker statement than '$\|A\|_S < 1$'. As a simple example, consider $A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ for some $\lambda \in (0,1)$. Then $\|A\|_S \geq 1+\lambda^2 > 1$.

6.7.1-13. Lemma. Let A be a real square asymptotically stable matrix and Γ an arbitrary square matrix of the same size as A. Then $\{A^k\Gamma(A^T)^k\}_{k=1}^{\infty}$ has e.d.

Proof. Consider $\|A^k\Gamma(A^T)^k\|_S \leq \|A^k\|_S^2\|\Gamma\|_S$ and apply the previous lemma.

Q.E.D.

6.7.1-14. Lemma. Let $\{A_k\}_{k=1}^{\infty}$ be a sequence of d×d real positive semi-definite symmetric matrices. Then $\{A_k\}_{k=1}^{\infty}$ is an e.d. sequence iff $\{\text{tr } A_k\}_{k=1}^{\infty}$ is an e.d. sequence.

Proof. If A is (real) positive semi-definite symmetric, its largest eigenvalue is $\|A\|_S$, and $\|A\|_S \leq \text{tr } A \leq d.\|A\|_S$. The lemma follows easily from this inequality.

Q.E.D.

6.7.1-15. Remarks. (i) Similar results hold for double matrix-sequences $\{A_{k\ell} | A_{k\ell}$ a $d_1 \times d_2$ matrix, $(k,\ell) \in \bar{\mathbb{N}}\}$.
(ii) Theorem (6.7.1-5) can be generalized to the matrix case in a straightforward manner.

For stochastic vectors let us introduce the concept of exponential decay of dependence.

6.7.1-16. Definition. Let $\{x_k\}_1^\infty$ be a sequence of random variables taking their values in B. It will be said to have the property of exponential decay of dependence if the following holds.
(i) For each $p \in \mathbb{N}$ the sequence $\{E\|x_k\|^p\}_{k=1}^\infty$ is bounded and
(ii) for each pair $(k,\ell) \in \bar{\mathbb{N}}$ there exist random variables $x_{k\ell}$ and $\varepsilon_{k\ell}$ such that
(a) $x_k = x_{k\ell} + \varepsilon_{k\ell}$,
(b) $x_{k\ell}$ is stochastically independent of $x_\ell, x_{\ell-1}, x_{\ell-2}, \ldots$ and
(c) $\forall p \in \mathbb{N}$: $\{E\|\varepsilon_{k\ell}\|^p | (k,\ell) \in \bar{\mathbb{N}}\}$ is an e.d. double sequence.

6.7.1-17. Notation. If $\{x_k\}_1^\infty$ has this property we will say that '$\{x_k\}_1^\infty$ has e.d.d.' or that '$\{x_k\}_{k=1}^\infty$ is an e.d.d. sequence'.

For those cases in which the $\varepsilon_{k\ell}$ are Gaussian, condition (c) of definition (6.7.1-16) simplifies considerably. To show this, the following lemma will be used.

6.7.1-18. Lemma. There exists a sequence of polynomials $\{q_p(\mu,\sigma^2)\}_{p=1}^\infty$ in the (scalar) variables μ and σ^2, with the following property. If x is a scalar Gaussian variable with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, then

(6.7.1-19) $Ex^p = q_p(\mu,\sigma^2)$.

Furthermore for each $p \in \mathbb{N}$, $q_p(0,0) = 0$.

Proof. This is a standard result from statistics. (It can easily be shown by making use of the characteristic function of the Gaussian distribution with

mean μ and variance $\sigma^2 \geq 0$).

Q.E.D.

6.7.1-20. Theorem. Let $\{x_k\}_{k=1}^{\infty}$ be a sequence of random variables taking their values in B. It has e.d.d. if

(i) for each $p \in \mathbb{N}$ the sequence $\{E\|x_k\|^p\}_{k=1}^{\infty}$ is bounded and

(ii) for each $(k,\ell) \in \bar{\mathbb{N}}$ there exist random variables $x_{k\ell}$ and $\varepsilon_{k\ell}$ such that

(a) $x_k = x_{k\ell} + \varepsilon_{k\ell}$,

(b) $x_{k\ell}$ is stochastically independent of $x_\ell, x_{\ell-1}, x_{\ell-2}, \ldots$,

(c) $\varepsilon_{k\ell}$ has a Gaussian distribution and

(d) $\{E\|\varepsilon_{k\ell}\|^2 \,|\, (k,\ell) \in \bar{\mathbb{N}}\}$ is an e.d. double sequence.

Proof. Let the mean and covariance matrix of $\varepsilon_{k\ell}$ be denoted by $\mu_{k\ell}$ and $\Sigma_{k\ell}$ respectively and let $\sigma^2_{k\ell} := \operatorname{tr} \Sigma_{k\ell}$. Because

$$\|\mu_{k\ell}\| \leq E\|\varepsilon_{k\ell}\| \leq \left(E\|\varepsilon_{k\ell}\|^2\right)^{\frac{1}{2}} \tag{6.7.1-21}$$

and

$$\sigma^2_{k\ell} = E\|\varepsilon_{k\ell}\|^2 - \|\mu_{k\ell}\|^2 \leq E\|\varepsilon_{k\ell}\|^2 \tag{6.7.1-22}$$

and $\{E\|\varepsilon_{k\ell}\|^2\}$ is an e.d. double sequence, it follows that $\{\mu_{k\ell}\}, \{\sigma^2_{k\ell}\}$ are e.d. double sequences. Using lemma (6.7.1-13) it follows that $\{\Sigma_{k\ell}\}$ is an e.d. double matrix sequence.

Because the conditions (i), (ii)(a) and (ii)(b) of definition (6.7.1-16) are assumed to be fulfilled, it remains to show (ii)(c) of definition (6.7.1-16), i.e. that for all $p \in \mathbb{N}$, $\{E\|\varepsilon_{k\ell}\|^p \,|\, (k,\ell) \in \bar{\mathbb{N}}\}$ is an e.d. double sequence. From the well-known inequality

$$(E\|\varepsilon_{k\ell}\|^j)^2 \leq E(\|\varepsilon_{k\ell}\|^{2j}) \tag{6.7.1-23}$$

it follows easily that it suffices to consider only even values of p. So let $p = 2j$, $j \in \mathbb{N}$. First consider the scalar case $d = 1$. Then $E\varepsilon^{2j}_{k\ell} = q_{2j}(\mu_{k\ell}, \sigma^2_{k\ell})$, according to lemma (6.7.1-18). Because $\{\mu_{k\ell}\}$ and $\{\sigma^2_{k\ell}\}$ have e.d., the same holds for $\{q_{2j}(\mu_{k\ell}, \sigma^2_{k\ell}) \,|\, (k,\ell) \in \bar{\mathbb{N}}\}$, for each $j \in \mathbb{N}$, according to theorem (6.7.1-5)(ii)(f).

Now consider the vector case. Use will be made of the following inequality. If

$x^{(1)},x^{(2)},\ldots,x^{(d)}$ are arbitrary scalar random variables then

$$(6.7.1\text{-}24)\quad E|x^{(1)}x^{(2)}\ldots x^{(d)}| \leq \prod_{1}^{d} [E(x^{(i)})^{2^{d-1}}]^{2^{1-d}}.$$

This inequality follows by induction from the well-known inequality for arbitrary scalar random variables a and b:

$$(6.7.1\text{-}25)\quad E|ab| \leq (Ea^2)^{\frac{1}{2}}(Eb^2)^{\frac{1}{2}}.$$

Let $\varepsilon_{k\ell}^{(i)}$ denote the i-th component of $\varepsilon_{k\ell}$, $i \in \{1,2,\ldots,d\}$.
One has for each $j \in \mathbb{N}$

$$(6.7.1\text{-}26)\quad E\|\varepsilon_{k\ell}\|^{2j} = E[\sum_{i=1}^{d} (\varepsilon_{k\ell}^{(i)})^2]^j =$$

$$\leq \sum_{\substack{0\leq j_i\leq j \\ \Sigma j_i = j}} \binom{j}{j_1,j_2,\ldots,j_d} \prod_{i=1}^{d} [E(\varepsilon_{k\ell}^{(i)})^{2^d}]^{2^{1-d}},$$

where (6.7.1-24) is used. From the scalar case it follows that for each $i \in \{1,2,\ldots,d\}$ the double sequence $\{E(\varepsilon_{k\ell}^{(i)})^{2^d} | (k,\ell) \in \bar{N}\}$ has e.d. It follows from theorem (6.7.1-5)(ii)(f) that the double sequence that is obtained by taking the right-hand side of (6.7.1-26) and letting (k,ℓ) take all possible values in $\bar{N}$, has e.d. And therefore $\{E\|\varepsilon_{k\ell}\|^{2j} | (k,\ell) \in \bar{N}\}$ has e.d.

Q.E.D.

It is perhaps not so surprising that the covariance matrix sequence of an e.d.d. sequence $\{x_k\}$ has e.d. This is what will be shown next.
Let $\mathrm{cov}(x_k,x_\ell) := E(x_k-Ex_k)(x_\ell-Ex_\ell)^T$. (As stated before we restrict ourselves to the real case, for the complex case similar results hold).

6.7.1-27. Theorem. Let $\{x_k\}_{k=1}^{\infty}$ be an e.d.d. sequence of vectors taking their values in B. The double sequence of covariance matrices
$\{\mathrm{cov}(x_k,x_\ell) | (k,\ell) \in \bar{N}\}$ has e.d.

Proof. For all $(k,\ell) \in \bar{N}$, one has

$$\text{cov}(x_k,x_\ell) = E(x_k-Ex_k)(x_\ell-Ex_\ell)^T =$$

(6.7.1-28)

$$= E(x_{k\ell}-Ex_{k\ell})(x_\ell-Ex_\ell)^T + E(\varepsilon_{k\ell}-E\varepsilon_{k\ell})(x_\ell-Ex_\ell)^T.$$

Because $x_{k\ell}$ and x_ℓ are stochastically independent they have vanishing covariances. It follows that

(6.7.1-29) $$\text{cov}(x_k,x_\ell) = E(\varepsilon_{k\ell}-E\varepsilon_{k\ell})(x_\ell-Ex_\ell)^T$$

Let us now take the Frobenius norm.

$$\|\text{cov}(x_k,x_\ell)\|_F = \|E(\varepsilon_{k\ell}-E_{k\ell})(x_\ell-Ex_\ell)^T\|_F \leq$$

(6.7.1-30)

$$E\|(\varepsilon_{k\ell}-E\varepsilon_{k\ell})(x_\ell-Ex_\ell)^T\|_F = E\{\|\varepsilon_{k\ell}-E\varepsilon_{k\ell}\|.\|x_\ell-Ex_\ell\|\} \leq$$

$$(E\|\varepsilon_{k\ell}-E\varepsilon_{k\ell}\|^2)^{\frac{1}{2}}(E\|x_\ell-Ex_\ell\|^2)^{\frac{1}{2}} \leq (E\|\varepsilon_{k\ell}\|^2)^{\frac{1}{2}}(E\|x_\ell\|^2)^{\frac{1}{2}}.$$

Now $\{E\|\varepsilon_{k\ell}\|^2\}_{k\ell}$ is an e.d. double sequence and $\{E\|x_\ell\|^2\}_\ell$ is a bounded sequence, so $\{(E\|\varepsilon_{k\ell}\|^2)^{\frac{1}{2}}(E\|x_\ell\|^2)^{\frac{1}{2}}\}_{k\ell}$ is an e.d. double sequence, hence $\{\|\text{cov}(x_k,x_\ell)\|_F \,|\, (k,\ell) \in \bar{N}\}$ is an e.d. double sequence and the theorem follows.

Q.E.D.

6.7.1-31. Corollary. If $\{x_k\}_1^\infty$ has e.d.d., then $\{\text{tr cov}(x_k,x_\ell) \,|\, (k,\ell) \in \bar{N}\}$ has e.d.

In many important cases, taking a function of an e.d.d. sequence gives another e.d.d. sequence. This is treated next.

6.7.1-32. Theorem. Suppose $\{x_k = (x_k^{(1)},x_k^{(2)},\ldots,x_k^{(d)})\}_{k=1}^\infty$ is an e.d.d. sequence of random vectors taking their values in B.
(a) Suppose $p = p(x^{(1)},x^{(2)},\ldots,x^{(d)})$ is a polynomial in d variables, then the sequence $\{p(x_k^{(1)},x_k^{(2)},\ldots,x_k^{(d)})\}_{k=1}^\infty$ has e.d.d.
(b) Consider $F: B \to \tilde{B}$, $\tilde{B}$ another finite dimensional vector space with inner product. Suppose F satisfies a global Lipschitz condition

(6.7.1-33) $$\exists c > 0 \quad \forall x,y \in B: \|F(x)-F(y)\| \leq c\|x-y\|.$$

Then $\{F(x_k)\}_{k=1}^{\infty}$ has e.d.d.

(c) If $F: B \to \tilde{B}$ is a C^1 mapping with compact support then $\{F(x_k)\}_{k=1}^{\infty}$ has e.d.d.

(d) If $\{F_k\}_{k=1}^{\infty}$ is a sequence of mappings satisfying a uniform Lipschitz condition

(6.7.1-34) $\exists c > 0,\ \forall x,y \in B, \forall k \in \mathbb{N}: \|F_k(x)-F_k(y)\| \leq c\|x-y\|.$

Then $\{F_k(x_k)\}_{k=1}^{\infty}$ has e.d.d.

(e) If $F: \tilde{B} \times B \to \tilde{B}$, $(\theta,x) \longmapsto F(\theta,x)$ is a continuous mapping with continuous partial derivatives with respect to the components of the x-vector, and F has compact support, then for each sequence $\{\theta_k \in \tilde{B}\}_{k=1}^{\infty}$, the sequence $\{F(\theta_k,x_k)\}_{k=1}^{\infty}$ has e.d.d.

(f) If $\lambda \in (0,1)$ then $\{y_k = \sum_{j=o}^{k-1} \lambda^j x_{k-j}\}_1^{\infty}$ has e.d.d.

(g) If $\{A_j\}_1^{\infty}$ is a sequence of linear mappings $A_j: B \to B$, and if a $\lambda \in (0,1)$ exists such that $\|A_j\| < \lambda$ for all j, then $\{y_k = \sum_{j=o}^{k-1} A_k A_{k-1} \cdots A_{k-j+1} x_{k-j}\}$ has e.d.d.

Proof. (a) First consider the case of two variables $x_k^T = (u_k,v_k)$, with $x_{k\ell}^T = (u_{k\ell},v_{k\ell})$ and $\varepsilon_{k\ell}^T = (\varepsilon_{k\ell}^u,\varepsilon_{k\ell}^v)$. The simplest cases are (i) $p(u,v) = u + v$ and (ii) $p(u,v) = uv$.

ad(i) Let $z_k = p(u_k,v_k) = u_k + v_k$. Let $z_{k\ell} := u_{k\ell} + v_{k\ell}$ and $\varepsilon_{k\ell}^z = \varepsilon_{k\ell}^u + \varepsilon_{k\ell}^v$. Then $z_{k\ell}$ is clearly stochastically independent of $z_\ell, z_{\ell-1}, z_{\ell-2}, \ldots$, and $\|\varepsilon_{k\ell}^z\| \leq \|\varepsilon_{k\ell}^u\| + \|\varepsilon_{k\ell}^v\|$ and so $\{\varepsilon_{k\ell}^z\}$ is an e.d. double sequence ($(k,\ell) \in \bar{\mathbb{N}}$) This shows (i).

ad(ii) Let $z_k = p(u_k,v_k) = u_k v_k$. Let $z_{k\ell} := u_{k\ell} v_{k\ell}$, then $z_{k\ell}$ is stochastically independent of $z_\ell, z_{\ell-1}, z_{\ell-2}, \ldots$. Let $\varepsilon_{k\ell}^z = u_k v_k - u_{k\ell} v_{k\ell}$.
Then one has

(6.7.1-35) $\varepsilon_{k\ell}^z = u_k v_k - (u_k-\varepsilon_{k\ell}^u)(v_k-\varepsilon_{k\ell}^v) = \varepsilon_{k\ell}^u v_k + \varepsilon_{k\ell}^v u_k - \varepsilon_{k\ell}^u \varepsilon_{k\ell}^v$

and so for each $p \in \mathbb{N}$

(6.7.1-36) $\|\varepsilon^z_{k\ell}\|^p \le \{|\varepsilon^u_{k\ell}||v_k| + |\varepsilon^v_{k\ell}||u_k| + |\varepsilon^u_{k\ell}||\varepsilon^v_{k\ell}|\}^p.$

The right-hand side is a polynomial in the three variables $|\varepsilon^u_{k\ell}||v_k|, |\varepsilon^v_{k\ell}||u_k|$ and $|\varepsilon^u_{k\ell}||\varepsilon^v_{k\ell}|$, without constant term. Using (6.7.1-24) one can find un upper estimate for $E\|\varepsilon^z_{k\ell}\|^p$ in terms of positive (but not necessarily integer) powers of moments of $|\varepsilon^u_{k\ell}||v_k|$, $|\varepsilon^v_{k\ell}||u_k|$ and $|\varepsilon^u_{k\ell}||\varepsilon^v_{k\ell}|$.
Similarly as is done in the proof of (6.7.1-20) one can show that for each p, $\{E\|\varepsilon^z_{k\ell}\|^p\}$ is an e.d. double sequence.
The arguments used for (i) and (ii) can be extended without problems to the case of d variables and to the case of polynomials p of any finite degree in those d variables (or one can use inductions). The details are left to the reader.
(b) Let $z_k := F(x_k)$ and $z_{k\ell} := F(x_{k\ell})$, then $\varepsilon^z_{k\ell} = z_k - z_{k\ell}$. Then $z_{k\ell}$ is independent of $z_\ell, z_{\ell-1}, z_{\ell-2}, \ldots$ and $\|\varepsilon^z_{k\ell}\| = \|F(x_k)-F(x_{k\ell})\| \le c\|x_k-x_{k\ell}\| = c\|\varepsilon_{k\ell}\|$.
This implies

$$E\|\varepsilon^z_{k\ell}\|^p \le c^p\, E\|\varepsilon_{k\ell}\|^p,$$

so $\{E\|\varepsilon^z_{k\ell}\|^p \mid (k,\ell) \in \bar{N}\}$ is an e.d. double sequence for each $p \in \mathbb{N}$ and so $\{z_k\}$ has e.d.d.
(c) F satisfies a Lipschitz condition, so (b) is applicable.
(d) Let $z_k = F(x_k)$ and $z_{k\ell} = F(x_{k\ell})$. The proof is now completely similar to the proof of (b).
(e) F satisfies a Lipschitz condition with respect to x that is independent of θ.
Therefore this case reduces to case(d).
(f) Let $y_{k\ell} := \sum_{j=o}^{k-\ell-1} \lambda^j x_{k-j,\ell}$, then $y_{k\ell}$ is stochastically independent of $x_\ell, x_{\ell-1}, \ldots$ and therefore of $y_\ell, y_{\ell-1}, y_{\ell-2}, \ldots$. One has

(6.7.1-37) $$\varepsilon^y_{k\ell} = y_k - y_{k\ell} = \sum_{j=o}^{k-\ell-1} \lambda^j \varepsilon_{k-j,\ell} + \sum_{j=k-\ell}^{k-1} \lambda^j x_{k-j}.$$

It follows that

$$(6.7.1\text{-}38)\quad \|\varepsilon^{y}_{k\ell}\| \leq \sum_{j=o}^{k-\ell-1} \lambda^{j}\|\varepsilon_{k-j,\ell}\| + \lambda^{k-\ell} \sum_{j=k-\ell}^{k-1} \lambda^{j-(k-\ell)}\|x_{k-j}\|.$$

Because $\{x_k\}$ has e.d.d., there exists a $c_1 > 0$ such that $\forall k \in \mathbb{N}$: $E\|x_k\| < c_1$ and there exists a $c_2 > 0$ and a $\lambda_2 \in (\lambda,1)$ such that $E\|\varepsilon_{k\ell}\| < c_2\lambda^{k-\ell}$ for all $(k,\ell) \in \bar{\mathbb{N}}$. Substituting this one obtains

$$(6.7.1\text{-}39)\quad E\|\varepsilon^{y}_{k\ell}\| \leq \sum_{j=o}^{k-\ell-1} \lambda_2^{j}\lambda_2^{k-j-\ell}c_2 + \lambda_2^{k-\ell} \sum_{i=o}^{\ell-1} \lambda_2^{i}c_1 \leq \lambda_2^{k-\ell}\{(k-\ell-1)c_2 + \frac{c_1}{1-\lambda_2}\}.$$

The right-hand side is the general term of a (k,ℓ)-double sequence, with $(k,\ell) \in \bar{\mathbb{N}}$, which has e.d., according to theorem (6.7.1-5) (ii) (c) and (a). It follows that $\{E\|\varepsilon^{y}_{k\ell}\| \mid (k,\ell) \in \bar{\mathbb{N}}\}$ has e.d.
Similarly as in the proof of (a) and of (6.7.1-20) this argument can be extended to show that for <u>each</u> $p \in \mathbb{N}$, $\{E\|\varepsilon^{y}_{k\ell}\|^{p} \mid (k,\ell) \in \bar{\mathbb{N}}\}$ has e.d., and so $\{y_k\}$ has e.d.d. The details are left to the reader.

(g) Let $y_{k\ell} := \sum_{j=o}^{k-\ell-1} A_k A_{k-1}\cdots A_{k-j+1}x_{k-j,\ell}$ then $y_{k\ell}$ is stochastically independent of $x_\ell, x_{\ell-1}, \ldots$ and therefore of $y_\ell, y_{\ell-1}, \ldots$. One has

$$(6.7.1\text{-}40)\quad \varepsilon^{y}_{k\ell} = y_k - y_{k\ell} = \sum_{j=o}^{k-\ell-1} A_k A_{k-1}\cdots A_{k-j+1}\varepsilon_{k-j,\ell} + \sum_{j=k-\ell}^{k-1} A_k\cdots A_{k-j+1}x_{k-j}.$$

Because for all j, $\|A_j\| < \lambda$ it follows that

$$(6.7.1\text{-}41)\quad \|\varepsilon^{y}_{k\ell}\| \leq \sum_{j=o}^{k-\ell-1} \lambda^{j}\|\varepsilon_{k-j,\ell}\| + \sum_{j=k-\ell}^{k-1} \lambda^{j}\|x_{k-j}\|.$$

This is equal to the inequality (6.7.1-37) with λ_2 instead of λ. From here the proof is identical to the proof of (f).

Q.E.D.

The previous theorem shows that the concept of an e.d.d. sequence of random variables is a rather flexible one. This will be useful in the analysis of the algorithm. However, a somewhat more general concept will in fact be needed, which is introduced next.

6.7.1-42. <u>Definition</u>. Let $\overset{\circ}{N} = \{N_i\}_{i=1}^{\infty}$ be a sequence of disjoint intervals of $\mathbb{N}$ with $N_i \leq N_{i+1}$ (i.e. $\forall x \in N_i, \forall y \in N_{i+1}, x \leq y$) and with interval lengths going to infinity for $i \to \infty$ and let

$N_o := \bigcup_{i=1}^{\infty} N_i$ and $\bar{N}_i := \{(k,\ell) | k \geq \ell, k \in N_i, \ell \in N_i\}$.

(a) Let $\{b_{k\ell} | (k,\ell) \in \cup \bar{N}_i\}$ be a deterministic double sequence of elements of B. Then $\{b_{k\ell} | (k,\ell) \in \cup \bar{N}_i\}$ is said to have the property of <u>exponential decay</u> with respect to the <u>interval sequence</u> $\overset{\circ}{N}$ (abbreviated as 'e.d.i-$\overset{\circ}{N}$', or 'e.d.i' if $\overset{\circ}{N}$ is clear from the context), if the following holds:

$\exists c > 0$, $\exists \lambda \in (0,1)$ such that $\forall (k,\ell) \in \cup \bar{N}_i$: $\|b_{k\ell}\| < c.\lambda^{k-\ell}$.

(b) Let $\{x_k | k \in N_o\}$ be a sequence of random variables taking their values in B. Then $\{x_k | k \in N_o\}$ is said to have the property of <u>exponential decay of dependence</u> with respect to the <u>interval sequence</u> $\overset{\circ}{N}$ (abbreviated as 'e.d.d.i-$\overset{\circ}{N}$' or 'e.d.d.i' if $\overset{\circ}{N}$ is clear from the context), if the following holds $\forall (k,\ell) \in \cup \bar{N}_i$, $\exists x_{k\ell}, \varepsilon_{k\ell}$ such that

(i) $x_k = x_{k\ell} + \varepsilon_{k\ell}$, $\forall (k,\ell) \in \bar{N}_i, \forall i \in \mathbb{N}$,
(ii) $x_{k\ell}$ is stochastically independent of $\{x_j | j \in N_i, j \leq \ell\}$, $\forall (k,\ell) \in \bar{N}_i$, $\forall i \in \mathbb{N}$, and
(iii) $\forall p \in \mathbb{N}, \exists c_p > 0$, $\exists \lambda_p \in (0,1)$ such that $\forall i \in \mathbb{N}$, $\forall (k,\ell) \in \bar{N}_i$: $E\|\varepsilon_{k\ell}\|^p < c_p \lambda_p^{k-\ell}$.

6.7.1-43. <u>Theorem</u>. Let $\{x_k | k \in N_o\}$ be an e.d.d.i.-$\overset{\circ}{N}$ sequence. The properties (a) - (g) of theorem (6.7.1-32) hold if

(i) 'e.d.d.' is replaced by 'e.d.d.i-$\overset{\circ}{N}$' <u>and</u>
(ii) in (f) and (g) the summation is replaced by summations that are going back only to the beginning of the relevant interval in $\overset{\circ}{N}$. I.e. in the analogon of (f)

$$\{y_k = \sum_{j=o}^{k-\min N_i} \lambda^j x_{k-j} | i \text{ such that } k \in N_i\}_{k \in N_o}$$

is the sequence that has e.d.d.i. and in the analogon of (g),

$$\{y_k = \sum_{j=o}^{k-\min N_i} A_k \ldots A_{k-j+1} x_{k-j} | i \text{ such that } k \in N_i\}_{k \in N_o}$$

is the sequence that has e.d.d.i.

The proof is completely analogous to the proof of theorem (6.7.1-32).

6.7.1-44. Lemma. If $\{x_k\}_1^\infty$ is an e.d.d. sequence, then $\{x_k\}_{k\in N_o}$ has e.d.d.i-$\overset{\circ}{N}$.

This is an obvious but useful result.

6.7.1-45. Lemma. If $\{x_k\}_{k\in N_o}$ is a sequence of random vectors such that

(i) $x_k = 0$ if, for some $i \in \mathbb{N}$, $k \in \mathbb{N}_i$ and $k \neq \min(N_i)$,
(ii) for each $p \in \mathbb{N}$ the sequence $\{E\|x_k\|^p\}_{k\in N_o}$ is bounded,
then $\{x_k\}_{k\in N_o}$ has e.d.d.i-$\overset{\circ}{N}$.

This can be shown simply by taking $x_{k\ell} = 0$ for all $(k,\ell) \in \cup_i \bar{N}_i$. The details are left to the reader.

6.7.2. Applications to the algorithm

First it will be shown that several quantities that appear in the algorithm have e.d.d. or e.d.d.i, especially $\{b_k\}$. This result will then be used to show that $\{f_k - \tilde{E}f_k\}$ converges to zero for $k \to \infty$, $\tilde{P}$-a.s. This implies that the hypothesis (6.6.3-5) is true.

6.7.2-1. Theorem. The sequence of outputs $\{y_t\}$ of the system to-be-identified is an e.d.d. sequence.

Proof. By assumption (cf. section 6.2.1) the 'true' model has an innovations representation (cf. section 2.4 and (4.8-1)) with $D = I$ (without loss of generality this may be assumed) and, for notational reasons, w_t instead of v_t:

$$(6.7.2\text{-}2)\qquad \begin{cases} x_{t+1} = Ax_t + Bw_t\,, & t \in \mathbb{Z},\ x_t \in \mathbb{R}^n, \\ y_t = Cx_t + w_t\,, & y_t \in \mathbb{R}^p,\ w_t \in \mathbb{R}^p \end{cases}$$

with A asymptotically stable etc. (cf. sections 2.4 and 4.8). For simplicity of notation assume in this proof that $t_o = 1$, and consider $\{y_t\}_{t=1}^\infty$. It is clear that

$$(6.7.2\text{-}3)\qquad y_t = CA^{t-1}x_1 + \sum_{r=1}^{t-1} CA^{r-1}Bw_{t-r} + w_t.$$

Now for $(t,s) \in \bar{N}$, let

$$(6.7.2\text{-}4) \qquad y_{ts} = \begin{cases} \sum_{r=1}^{t-s-1} CA^{r-1}Bw_{t-r} + w_t & \text{if } t > s, \\ 0 & \text{otherwise.} \end{cases}$$

Then y_{ts} is stochastically independent of $w_s, w_{s-1}, w_{s-2}, \ldots$ and therefore of $y_s, y_{s-1}, y_{s-2}, \ldots$.
Let

$$(6.7.2\text{-}5) \qquad \varepsilon^y_{ts} = y_t - y_{ts} = CA^{t-1}x_1 + \sum_{r=t-s}^{t-1} CA^{r-1}Bw_{t-r}.$$

It follows that

$$(6.7.2\text{-}6) \qquad E\varepsilon_{ts}\varepsilon_{ts}^T = CA^{t-1}Ex_1x_1^T(A^{t-1})^TC^T + \sum_{r=t-s}^{t-1} CA^{r-1}B\,\Sigma\,B^T(A^{r-1})^TC^T.$$

From lemma (6.7.1-13) it follows that $\{CA^{r-1}B\Sigma B^T(A^{r-1})^TC^T\}_{r=1}^{\infty}$ is e.d. Let the sequence $\{c_r\}_1^{\infty}$ be defined by

$$(6.7.2\text{-}7) \qquad c_r = \operatorname{tr} CA^{r-1}B\,\Sigma\,B^T(A^{r-1})^TC^T, \quad r = 1,2,3,\ldots .$$

Then $\{c_r\}$ has e.d.; this follows from lemma (6.7.1-13) and theorem (6.7.1-5) (i) (a), making use of the fact that f_r is a linear combination of the entries of $A^{r-1}B\,\Sigma\,B^T(A^T)^{r-1}$.
It follows that

$$\left\{ \sum_{r=t-s}^{t-1} c_r \right\}_{(t,s)\in\bar{N}}$$

is an e.d. double sequence, as can easily be shown. The sequence $\{\operatorname{tr} CA^{t-1}Ex_1x_1^T(A^{t-1})^TC^T\}_{t=1}^{\infty}$ is e.d. (same proof as for the fact that $\{c_r\}$ has e.d.). It follows easily that $\{E\|\varepsilon_{ts}\|^2\} = \{\operatorname{tr} E\varepsilon_{ts}\varepsilon_{ts}^T\}$ is an e.d. double sequence. Because the ε_{ts} are all Gaussian, theorem (6.7.1-20) tells us that $\{y_t\}$ has e.d.d.

Q.E.D.

6.7.2-8. Corollary. $\{\|y_t\|\}$ has e.d.d.

Proof. $F(y) = \|y\|$ is mapping with Lipschitz constant 1:

(6.7.2-9) $\quad |F(y)-F(x)| = |\|y\| - \|x\|| \leq \|y-x\|.$

Therefore theorem (6.7.1-32) (b) is applicable and the corollary follows.

Q.E.D.

6.7.2-10. Theorem. $\{v_t\}$ has e.d.d.

Proof. Let $t_o = 1$ to simplify the notation (without loss of generality). From the definition (6.2.8-3) of v_t it follows that

(6.7.2-11) $\quad v_t = \sum_{r=o}^{t-1} \lambda_1^r \|y_{t-r}\|.$

Corollary (6.7.2-8) and theorem (6.7.1-32) can now be applied to conclude that $\{v_t\}$ has e.d.d.

Q.E.D.

Next we want to show that $\{g_1(v_t)\}$ has e.d.d. To be able to do this we need the following technical lemma.

6.7.2-12. Lemma. Let for all $(t,s) \in \bar{N}$

$$v_{ts} = \begin{cases} \sum_{r=o}^{t-s-1} \lambda_1^r \|y_{t-r,s}\| & \text{if } t > s, \\ 0 & \text{if } t = s. \end{cases}$$

The following holds: let $v > 0$, then

(6.7.2-13) $\quad \exists c > 0, \exists \varepsilon_o > 0, \forall (s,t) \in \bar{N}, \forall \varepsilon \in (0,\varepsilon_o)$: $P(v-\varepsilon < v_{st} \leq v) < c.\varepsilon^{\frac{1}{2}}.$

The proof will be given in appendix 6B.

6.7.2-14. Theorem. $\{g_1(v_t)\}$ has e.d.d.

Proof. Let

(6.7.2-15) $\varepsilon^{g}_{ts} := g_1(v_t) - g_1(v_{ts})$,

where v_{ts} is as defined in lemma (6.7.2-12). Clearly v_{ts} and therefore $g_1(v_{ts})$ is stochastically independent of $y_s, y_{s-1}, \ldots$ and therefore of $v_s, v_{s-1}, \ldots$ and of $g_1(v_s), g_1(v_{s-1}), \ldots$. It is clear that $\varepsilon^{g}_{ts} \in \{-1,0,1\}$. One has (cf. (6.2.8-4) for the definition of g_1)

$$\forall p \in \mathbb{N}: E|\varepsilon^{g}_{ts}|^p = E|\varepsilon^{g}_{ts}| = P[g_1(v_t) \neq g_1(v_{ts})] =$$
$$= P[v_t > K' \text{ and } v_{ts} \leq K'] + P[v_t \leq K' \text{ and } v_{ts} > K'] \tag{6.7.2-16}$$

Let $\varepsilon^{v}_{ts} = v_t - v_{ts}$. Then

$$P[v_t > K' \text{ and } v_{ts} \leq K'] = P[\varepsilon^{v}_{ts} > K' - v_{ts} \text{ and } v_{ts} \leq K'] \tag{6.7.2-17}$$

Let $F(v_{ts})$ denote the distribution function of v_{ts}, then

$$P[\varepsilon^{v}_{ts} > K' - v_{ts} \text{ and } v_{ts} \leq K'] = \int_{v_{ts}=0}^{K'} P[\varepsilon^{v}_{ts} > K' - v_{ts} | v_{ts}] dF(v_{ts}) = \tag{6.7.2-18}$$

$$= \int_{v_{ts}=0}^{K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}}} \frac{E(|\varepsilon^{v}_{ts}||v_{ts})}{K'-v_{ts}} dF(v_{ts}) + \int_{v_{ts}=K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}}}^{K'} dF(v_{ts}) \leq$$
$$\leq \int_{0}^{K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}}} \frac{E(|\varepsilon^{v}_{ts}||v_{ts})}{(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}}} dF(v_{ts}) + P[K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}} \leq v_{ts} \leq K']$$
$$\leq (E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}} + P[K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}} \leq v_{ts} \leq K'].$$

According to theorem (6.7.2-10) $\{v_t\}$ has e.d.d. and therefore $\{E|\varepsilon^{v}_{ts}|\}_{(t,s)\in\bar{N}}$ is an e.d. double sequence. It follows that

$\{(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}}\}$ is an e.d. double sequence too. And making use of lemma (6.7.2-12) it follows that $\{P[K'-(E|\varepsilon^{v}_{ts}|)^{\frac{1}{2}} \leq v_{ts} \leq K']\}_{(t,s)\in\bar{N}}$ is an e.d. double sequence. Combining (6.7.2-17) and (6.7.2-18), it follows that $\{P[v_t > K' \text{ and } v_{ts} \leq K'] | (t,s) \in \bar{N}\}$ is an e.d. double sequence. In a completely similar fashion, by interchanging the roles of v_t and v_{ts} and

replacing ε^v_{ts} by $-\varepsilon^v_{ts}$ one can show that $\{P[v_t \leq K' \text{ and } v_{ts} > K'] | (t,s) \in \bar{N}\}$ is an e.d. double sequence. Equation (6.7.2-16) now implies that for each $p \in \mathbb{N}$, $\{E|\varepsilon^g_{ts}|^p\}$ is an e.d. double sequence, and therefore $\{g_1(v_t)\}$ is e.d.d.

Q.E.D.

Now consider the decoupled algorithm, and let $\{\theta_t\}$ satisfy property 3 with interpolation curve $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta^+_t = Y^+(s_t), \forall t \in \mathbb{Z}$. Let $\tilde{S} \in \overset{\circ}{S}, \tilde{S} = (\{r_k\}_1^\infty, q, \delta)$, be such that $|I(\tilde{S}, Y^+)| = 1$, say $I(\tilde{S}, Y^+) = \{(i,j)\}$ and let $\overset{\circ}{N} = \{N_k\}_{k=1}^\infty$ be the corresponding sequence of intervals defined by

$$N_k = \{t | s_t \in [r_k - q, r_k]\}.$$

6.7.2-19. Theorem. $\{\xi_t\}$ has e.d.d.i.-$\overset{\circ}{N}$ (with respect to $\tilde{P}$).

Proof. Combining the inequality (6.4-12) of lemma (6.4-8) with lemma (6.4-41) (a) one finds there exists a $c_3 > 0$, c_3 data-independent, such that

$$\forall t \geq t_o: \|\xi_t\| \leq c_3 v_t.$$

Because $\{v_t\}$ has e.d.d. (theorem (6.7.2-10)), it follows that $\{\tilde{E}\|\xi_t\|^p\}_{t=t_o}^\infty$ is bounded for each $p \in \mathbb{N}$.

Now, according to theorem (6.3.4-21) one can choose (and we assume that this has been taken care of) the refinement of the coordinate-charts-cover of the manifold such that within each chart D_{ij} the spectral norm of $F(\phi_j(\theta);i,j)$ is smaller than $\lambda_o + \varepsilon (< 1)$. Now consider equation (6.3.3-15). If $t+1$, $t \in N_k$, then $\xi(t+1;i,j)$ can be decomposed as the sum of a vector due to $\xi(\min N_k;i,j)$ and the rest, which depends only on $y_t, y_{t-1}, \ldots, y_{\min(N_k)}$. The first part has e.d.d.i.-$\overset{\circ}{N}$ according to the combination of lemma (6.7.1-45) (with $x_t = \xi(t;i,j)$ if $t = \min N_k$ for some k and $x_t = 0$ otherwise) with theorem (6.7.1-43) (g) (with $F(\phi_j(\theta_t);i,j)$ instead of A_t). Because $\{y_t\}$ has e.d.d., $\{G(\phi_j(\theta_t);i,j)\}$ has e.d.d.i.-$\overset{\circ}{N}$ and therefore the second part has e.d.d.i.-$\overset{\circ}{N}$ according to theorem (6.7.1-43) (g) (again with $F(\phi_j(\theta_t);i,j)$ instead of A_t). Applying (6.7.1-43) (a) one finds that $\{\xi_t\}$ has

e.d.d.i.-$\overset{\circ}{N}$ (w.r.t.$\tilde{P}$).

Q.E.D.

6.7.2-20. Theorem. $\{b_t\}$ has e.d.d.i.-$\overset{\circ}{N}$ (w.r.t. $\tilde{P}$).

Proof. Consider the definition of b_t in (6.6.2-1). Substituting for $h(\xi_t,j)$ according to (6.2.8-5), one finds

(6.7.2-21) $\quad b_t = g_1(v_t)g_2(t)R(\theta_t,j)^{-1}\Psi(t,j)^T e(t,j).$

Consider (6.3.3-16); the matrix $H(\phi_j(\theta);i,j)$ is bounded on $\bar{D}'_{ij}$; $\{\xi(t;i,j)\}$ has e.d.d.i.-$\overset{\circ}{N}$ (theorem (6.7.2-19)) and $\{y_t\}$ has e.d.d. (theorem (6.7.2-1)). Therefore $\{z(t,j)\}$ has e.d.d.i.-$\overset{\circ}{N}$, where

$$z(t,j)^T = (e(t,j)^T,\Psi_1(t,j)^T,\ldots,\Psi_d(t,j)),$$

as before. Applying theorem (6.7.1-43) (a) (or in fact a slight generalization of this, namely to the case of a vector of polynomials instead of one polynomial) one finds that $\{\Psi(t,j)^T e(t,j)\}$ has e.d.d.i.-$\overset{\circ}{N}$. From corollary (6.4-55) it follows easily that $\{g_2(t)\}$ has e.d.d. Theorem (6.7.2-14) states that $\{g_1(v_t)\}$ has e.d.d. And $\{R(\theta_t,j_t)^{-1}\}$ is a sequence of bounded nonstochastic matrices. It follows from theorem (6.7.1-43) (a) that $\{b_t\}$ has e.d.d.i.-$\overset{\circ}{N}$.

Q.E.D.

6.7.2-22. Corollary. The double sequence

$$\{\tilde{E}(b_t-\tilde{E}b_t)^T(b_s-\tilde{E}b_s)\,|(t,s) \in \cup\, \bar{N}_i\} \text{ has e.d.i-}\overset{\circ}{N},$$

i.e. $\exists c > 0$, $\exists \lambda \in (0,1)$ such
that $\forall(t,s) \in \cup\, \bar{N}_i : |\tilde{E}(b_t-\tilde{E}b_t)^T(b_s-\tilde{E}b_s)| < c.\lambda^{t-s}$.

The proof is analogous to that of corollary (6.7.1-31), the only difference is that here it concerns 'e.d.d.i.-$\overset{\circ}{N}$' instead of instead of 'e.d.d.', and 'e.d.i' instead of 'e.d.'. The details are left to the reader. From this corollary the following important lemma can be derived. (Let $\{\theta_t\}$, $\tilde{S}$, $\overset{\circ}{N}$ etc. be again as

described just before theorem (6.7.2-19)).

6.7.2-23. Lemma. $\sum_{k=1}^{\infty} \tilde{E}\|f_k-\tilde{E}f_k\|^2 < \infty$.

Proof. Consider the definition (6.6.2-37) of f_k. Because $|I(\tilde{S},Y^+)| = 1$ it follows that $|I_k| = 1$ for each $k \in \mathbb{N}$; in fact $I_k = \{(i,j)\}$ for each k.
So for each $k \in \mathbb{N}$:

$$f_k = \int_{r_k-q+\delta}^{r_k} b^o(r)dr.$$

It follows that

$$(6.7.2\text{-}24)\quad f_k-\tilde{E}f_k = \int_{r_k-q+\delta}^{r_k} \{b^o(\sigma) - \tilde{E}b^o(\sigma)\}d\sigma$$

and

$$(6.7.2\text{-}25)\quad \sum_{k=1}^{\infty} \tilde{E}(f_k-\tilde{E}f_k)^T(f_k-\tilde{E}f_k) =$$

$$\sum_{k=1}^{\infty} \int_{r_k-q+\delta}^{r_k} \int_{r_k-q+\delta}^{r_k} \tilde{E}\{b^o(\sigma_1)-\tilde{E}b^o(\sigma_1)\}^T\{b^o(\sigma_2)-\tilde{E}b^o(\sigma_2)\}d\sigma_1 d\sigma_2.$$

Consider the definition (6.6.2-27) of $b^o(\sigma)$. Using the previous corollary one can derive

(6.7.2-26)

$$\int_{r_k-q+\delta}^{r_k} \int_{r_k-q+\delta}^{r_k} \tilde{E}\{b^o(\sigma_1)-\tilde{E}b^o(\sigma_1)\}^T\{b^o(\sigma_2)-\tilde{E}b^o(\sigma_2)\}d\sigma_1 d\sigma_2 \leq \sum_{(t,s)\in N_k^2} a_t a_s c\lambda^{|t-s|},$$

with $c > 0$, $\lambda \in (0,1)$ as in corollary (6.7.2-22).
A fortiori, it follows that

$$(6.7.2\text{-}27)\quad \tilde{E}\|f_k-\tilde{E}f_k\|^2 \leq c \sum_{\substack{t\geq t_o \\ s\geq t_o}} a_t a_s \lambda^{|t-s|}.$$

For simplicity of notation, take $t_o = 1$. Now consider

$$(6.7.2\text{-}28)\quad c\sum_{\substack{t\geq 1\\ s\geq 1}} a_t a_s \lambda^{|t-s|} \leq 2c \sum_{t\geq s\geq 1} a_t a_s \lambda^{t-s} \leq 2c \sum_{j=0}^{\infty} \Big(\sum_{s=1}^{\infty} a_{s+j} a_s \Big) \lambda^j \leq$$

$$\leq 2c \sum_{j=0}^{\infty} \Big(\sum_{s=1}^{\infty} a_s^2 \Big) \lambda^j = \frac{2c}{1-\lambda} \sum_{s=1}^{\infty} a_s^2 < \infty.$$

The lemma follows.

Q.E.D.

To show that from this it follows that the hypothesis (6.6.3-5) is true the following lemma is needed.

6.7.2-29. Lemma. Let $\{x_k\}_{k=1}^{\infty}$ be an arbitrary sequence of random vectors $x_k \in B$ with mean zero and covariance matrix Γ_k. If $\sum_{k=1}^{\infty} \operatorname{tr} \Gamma_k < \infty$, then $\lim_{k\to\infty} x_k = 0$ with probability one.

Proof. Let $\varepsilon > 0$ and apply Chebyshev's inequality

$$(6.7.2\text{-}30)\quad P(\|x_k\| > \varepsilon) \leq \frac{E x_k^T x_k}{\varepsilon^2} = \frac{\operatorname{tr} \Gamma_k}{\varepsilon^2}.$$

Therefore

$$(6.7.2\text{-}31)\quad \forall \varepsilon > 0: \sum_{k=1}^{\infty} P(\|x_k\| > \varepsilon) \leq \sum_{k=1}^{\infty} \operatorname{tr} \Gamma_k / \varepsilon^2 < \infty.$$

According to the lemma of Borel-Cantelli it follows that $\lim_{k\to\infty} x_k = 0$ with probability one.

Q.E.D.

Combining lemmas (6.7.2-23) and (6.7.2-29) one finds the main result of this section.

6.7.2-32. Theorem. The hypothesis (6.6.3-5) is true.

6.7.2-33. Remark. It follows of course that those results in section 6.6 (cf. especially theorem (6.6.3-16) and (6.6.3-17)) that are derived under the condition that the hypothesis is true, are all true!

6.8. The associated differential equation

6.8.1. An integral formula for the decoupled algorithm

In the second part of this section (section 6.8.2) the ordinary differential equation (o.d.e) that is associated with the algorithm will be derived. The construction of the function that satisfies the o.d.e. was an idea of [Ku-Cl]. However, in section 6.9 it will turn out that this function is constant, and its value is in fact an equilibrium point of the o.d.e. Therefore any claim that the algorithm will eventually follow or approximate a nonconstant solution curve of the differential equation seems unjustified, or is at least not justified by the approach followed here.
In this section (6.8.1) an integral formula will be derived for the decoupled algorithm, which is needed for the derivation of the o.d.e. in section 6.8.2. To start with, two sorts of variables that depend on a parameter point (instead of a parameter curve or sequence) will be defined.

6.8.1-1. Notation. Let $\theta^+ = (\theta,i,j) \in \Theta^+$ and consider the decoupled algorithm with constant parameter sequence $\{\theta_t^+ = \theta^+\}_{t=t_o}^{\infty}$. Then in the notation of the variables occurring in the decoupled algorithm, $\{\theta_t^+\}_{t=t_o}^{\infty}$ will be replaced by θ^+. For example, $b_t(\theta^+) = b_t(\{\theta^+\}_{t=t_o}^{\infty})$.
Now consider the vector of all random variables that occur in the decoupled algorithm at time t, or are derived from it. It is

$$(6.8.1\text{-}2) \qquad (y_t, v_t, g_1(v_t), \xi(t+1;i,j), e(t,j), \Psi(t,j), h, b_t).$$

If the parameter sequence has constant value θ^+, then the probability distribution of this vector converges to a steady state distribution for $t \to \infty$. (From corollary (6.4-55) it follows that $g_2(t) = 1$ for t large enough and it can be considered as nonrandom for $t \to \infty$. Therefore it can be left out of our considerations concerning the asymptotic behaviour of the algorithm).

6.8.1-3. Definition. Let $(y,v,g_1(v),\xi(\theta^+),\varepsilon(\theta),\Psi(\theta,j),h(\theta,j),b(\theta,j))$ be a random vector with as its probability distribution the steady state distribution of (6.8.1-2). The components of the vector will be called steady state random variables for the decoupled algorithm at the parameter point θ^+.

Note that $\varepsilon(\theta)$ does not depend on j. That this is correct follows from the

construction of the prediction error algorithm (cf. section 6.2, esp. section 6.2.2); $\varepsilon(\theta)$ is the steady state random variable corresponding the prediction error $\varepsilon_t(\theta)$ defined in (6.2.2-3). To be able to work with local coordinates, instead of $\varepsilon(\theta)$ also the notation $\varepsilon(\phi;j)$ will be used, where $\phi = \phi_j(\theta)$ stands for the local coordinates of θ in C_j (as before). Because the random variables in the definition correspond to the decoupled algorithm their expectation will be denoted by $\tilde{E}\xi(\theta^+)$, $\tilde{E}\varepsilon(\phi;j)$ etc.

6.8.1-4. Remarks. (i) One way to construct formally a steady state random vector for the decoupled algorithm is by taking a constant parameter sequence in the algorithm and starting formally at $t = -\infty$, which makes it into an asymptotically stable time-invariant filter. Because the true system is asymptotically stable, the result is mathematically well-defined and the resulting variables have the steady-state probability distribution at each time t.
(ii) If $\theta^+ = (\theta,i,j)$ then $b(\theta^+) = b(\theta,i,j)$ will denote the same as $b(\theta,j)$, this also holds for h, Ψ,R etc. The algorithm is constructed such that the following equality holds (cf. section 6.2 especially section 6.2.2).

6.8.1-5. Theorem. Let $\theta^+ = (\theta,i,j) \in \Theta^+$. One has

$$\tilde{E}b(\theta,j) = -R(\theta,j)^{-1} \frac{\partial}{\partial \phi} \tilde{E}g_1(v)\varepsilon(\phi;j)^T\varepsilon(\phi;j)/2, \tag{6.8.1-6}$$

i.e. $\tilde{E}b(\theta,j)$ equals minus the Riemannian gradient of the function

$$V_g(\theta) := \tfrac{1}{2}\tilde{E}g_1(v)\varepsilon(\theta)^T\varepsilon(\theta), \tag{6.8.1-7}$$

in terms of the local coordinates of the chart (C_j,ϕ_j).
If K' in the definition of g_1 is taken large enough, the probability that $g_1(v)$ equals zero will be very small, and $V_g(\theta)$ will be close to $V(\theta)$. There is an asymptotic relation between the value of b_k for a varying parameter sequence and its value for a related constant parameter sequence. The same holds for $\tilde{E}b_k$. But asymptotically b_k with constant parameter sequence has the same distribution as b and thus one obtains an asymptotic relation between $\tilde{E}b_k$ with a (specific type of) varying parameter sequence and $\tilde{E}b$. The precise formulation, in terms of interpolation curves, is as follows.

6.8.1-8. Theorem. Let the coordinate-chart index (i,j) be fixed and let (q,δ) with $q > \delta > 0$, be fixed. (a) For each $\varepsilon > 0$ there exists a t_2 such that if $Y^+ \in \mathring{L}^+_{cP}(-\infty,0]$ and $Y^+|_{[-q,0]} \in \mathring{L}^+_{ij}[-q,0]$ then

$$\forall t \geq t_2, \forall \omega \in \Omega, \forall \sigma \in [0,q-\delta], \|b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+) - b_t(Y^+(-\sigma))\| < \varepsilon,$$

where by abuse of notation $Y^+(-\sigma)$ denotes the parameter sequence with constant value $Y^+(-\sigma)$ (cf. (6.8.1-1)).

(b) For each $\varepsilon > 0$ there exists a t_5 such that if $Y^+ \in \mathring{L}^+_{cP}(-\infty,0]$ and $Y^+|_{[-q,0]} \in \mathring{L}^+_{ij}[-q,0]$, then

$$\forall t \geq t_5, \forall \sigma \in [0,q-\delta], \|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+) - \tilde{E}b(Y^+(-\sigma))\| < \varepsilon.$$

(c) Let $\{Y_t\}_{t=t_o}^{\infty}$ be a sequence such that

(i) $\forall t \geq t_o$: $Y^+_t \in \mathring{L}^+_{cP}(-\infty,0]$,

(ii) $\forall t \geq t_o$: $Y^+_t|_{[-q,0]} \in \mathring{L}^+_{ij}[-q,0]$ and

(iii) $\lim_{t\to\infty}(Y^+_t|_{[-q,0]}) = Y^+|_{[-q,0]}$ (convergence in the topology of $\mathring{L}^+_{ij}[-q,0]$, cf. section 6.5).

Then for each $\varepsilon > 0$ there exists a t_7 such that

(6.8.1-9) $\forall t \geq t_7, \forall \sigma \in [0,q-\delta]$: $\|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+_t) - \tilde{E}b(Y^+(-\sigma))\| < \varepsilon$.

Let $\{\tau(t)\}_{t=t_o}^{\infty}$ be a sequence of nonnegative numbers with limit zero. Then for each $\varepsilon > 0$ there exists a t_8 such that

(6.8.1-10) $\forall t \geq t_8, \forall \sigma \in [0,q-\sigma]$: $\|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma-\tau(t)}(Y^+_t) - \tilde{E}b(Y^+(-\sigma))\| < \varepsilon$.

Proof. (a) Because of lemma (6.6.2-3) it is sufficient to show that for each $\delta' > 0$ there exists a t_2 such that if $Y^+ \in \mathring{L}^+_{cP}(-\infty,0]$ and $Y^+|_{[-q,0]} \in \mathring{L}^+_{ij}[-q,0]$, then

$$\forall t \geq t_2, \forall \sigma \in [0, q-\delta]: \rho^+(\Pi_t^a \circ S_{-\sigma}(Y^+), \{Y(-\sigma)\}) < \delta'.$$

This property can be shown as follows. Let $n_o \in \mathbb{N}$ be such that $n_o^{-1} < \min(\delta', 1)$. Let t_2 be such that $\forall t \geq t_2$: $\sum_{s=t-n_o}^{t-1} a_s < \min(\delta, \frac{\delta'}{c})$ (where c is the Lipschitz constant!). From the definition of ρ^+ (cf. (6.5-12)) it follows that

$$\rho(\Pi_t^a \circ S_{-\sigma}(Y^+), \{Y^+(-\sigma)\}) \leq$$

(6.8.1-11)

$$\leq \max\{\sup_{0 \leq n \leq n_o-1} [\sup_{1 \leq j \leq n+1} d^+(Y^+(-\sum_{s=t-j}^{t-1} a_s - \sigma), Y^+(-\sigma))]; \sup_{n \geq n_o} (\tfrac{1}{n+1})\} \leq$$

$$\leq \max\{c \sum_{s=t-n_o}^{t-1} a_s, \frac{1}{n_o+1}\} < \max\{\delta, \delta\} = \delta.$$

(b) Because

$$\|\tilde{E}(b_t \circ \Pi_t^a \circ S_{-\sigma}(Y^+) - b_t(Y^+(-\sigma)))\| \leq \tilde{E}\|b_t \circ \Pi_t^a \circ S_{-\sigma}(Y^+) - b_t(Y^+(-\sigma))\|,$$

it follows from (a) that for each $\varepsilon > 0$ there exists a t_2 such that if $Y^+ \in \overset{\circ}{L}{}^+_{cP}(-\infty, 0]$ and $Y^+|_{[-q,0]} \in \overset{\circ}{L}{}^+_{ij}[-q, 0]$,

(6.8.1-12) $\forall t \geq t_2, \forall \sigma \in [0, q-\delta]: \|\tilde{E}b_t \circ \Pi_t^a \circ S_{-\sigma}(Y^+) - \tilde{E}b_t(Y^+(-\sigma))\| < \varepsilon/2.$

Therefore it suffices to show that for given $\varepsilon > 0$ there exists a t_6 such that

(6.8.1-13) $\forall t \geq t_6, \forall \theta^+ \in \bar{D}'^+_{ij}: \|\tilde{E}b_t(\theta^+) - \tilde{E}b(\theta^+)\| < \varepsilon/2.$

Once this has been shown then taking $\theta^+ = Y^+(0)$, $t_5 = \max(t_2, t_6)$ and combining (6.8.1-10) and (6.8.1-11) the result follows. Now consider (6.8.1-11). According to remark (6.8.1-4), $b(\theta^+)$ may be taken to be b_t^* if b_t^* is defined as the b_t that results if the algorithm is started (formally) at $t = -\infty$ (instead of $t = t_o$) and a constant parameter sequence $\{\theta^+\}$ is employed. The difference of b_t^* with the b_t that results if the algorithm is started at $t = t_o$ and the same constant parameter sequence $\{\theta^+\}$ is used, is caused (only) by the fact

that $\xi^*_{t_o}$ has a nonzero value while $\xi_{t_o} = 0$. (The notation is obvious: ξ^*_t is the vector $\xi(t;i,j)$ that results if in the decoupled algorithm starting at $t = -\infty$ the constant parameter sequence $\{\theta^+_t\}$ is applied, ξ_t is the vector $\xi(t;i,j)$ that results if in the decoupled algorithm starting at $t = t_o$ the constant parameter sequence $\{\theta^+\}$ is applied). The difference $\Delta\xi_t = \xi^*_t - \xi_t$ is equal to $F^{t-t_o}\Delta\xi_{t_o} = F^{t-t_o}\xi^*_{t_o}$. Using the fact that $\|F\|_S$ is bounded by a number smaller than one, uniformly for all $\theta^+ \in \bar{D}'^+_{ij}$ (cf. theorem (6.3.4-21)) it follows that $\tilde{E}(b^*_t - b_t)$ has exponential decay, uniformly for all $\theta^+ \in \bar{D}'^+_{ij}$, and the result follows.

(c) First (6.8.1-9) will be shown. Consider the following inequality

$$\begin{aligned} &\|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+_t) - \tilde{E}b(Y^+(-\sigma))\| \leq \\ &\leq \|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+_t) - \tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+)\| + \\ &+ \|\tilde{E}b_t \circ \Pi^a_t \circ S_{-\sigma}(Y^+_t) - \tilde{E}b(Y^+(-\sigma))\|. \end{aligned} \tag{6.8.1-14}$$

Let us consider separately each of the two terms in the right-hand side starting with the first. Let $\varepsilon > 0$ be given. According to lemma (6.6.2-3) there exists a $\delta' > 0$ such that

$$\forall\omega \in \Omega, \forall t \geq t_3, \forall\alpha,\beta \in \Pi^a_t(\mathring{L}^+_{cP}(-\infty,0]):$$

$$\rho^+(\alpha,\beta) < \delta' \Rightarrow \|b_{t\omega}(\alpha) - b_{t\omega}(\beta)\| < \varepsilon/2,$$

and so, a fortiori

$$\begin{aligned} &\forall t \geq t_3, \forall\alpha,\beta \in \Pi^a_t(\mathring{L}^+_{cP}(-\infty,0]): \\ &\rho^+(\alpha,\beta) < \delta' \Rightarrow \|\tilde{E}b_t(\alpha) - \tilde{E}b_t(\beta)\| < \varepsilon/2. \end{aligned} \tag{6.8.1-15}$$

Because $\lim_{t\to\infty}(Y_t^+|_{[-q,0]}) = Y^+|_{[-q,0]}$, there exists a $t' \geq t_o$ such that for all $t \geq t'$, $\sup_{-q\leq r\leq 0} d^+(Y_t^+(r),Y^+(r)) < \delta'$ and therefore

$$\forall \sigma \in [-q+\delta,0]: \sup_{-\delta\leq r\leq 0} d^+(Y_t^+(r-\sigma),Y^+(r-\sigma)) < \delta'.$$

Now consider proposition (6.5-14) (b) and the $n_o(t)$ defined there. Because $\lim_{t\to\infty} a_t = 0$, one has $\lim_{t\to\infty} n_o(t) = \infty$, and so there exists a $t'' \geq t_o$ such that $\forall t \geq t'': \frac{1}{n_o(t)+1} < \delta'$. Application of proposition (6.5-14) (b) gives:

(6.8.1-16) $\forall t \geq \max(t',t''): \rho^+(\Pi_t^a(S_{-\sigma}(Y_t^+)),\Pi_t^a(S_{-\sigma}(Y^+))) < \delta'.$

Combination of (6.8.1-15) with (6.8.1-16) gives:

(6.8.1-17)

$$\forall t \geq \max(t',t'',t_3),\forall \sigma \in [-q+\delta,0]: \|\tilde{E}b_t \circ \Pi_t^a \circ S_{-\sigma}(Y_t^+) - \tilde{E}b_t \circ \Pi_t^a \circ S_\sigma(Y^+)\| < \frac{\varepsilon}{2}$$

According to (b) there exists a t_5 such that

(6.8.1-18) $\forall t \geq t_5,\forall\sigma\in[-q+\delta,0]: \|\tilde{E}b_t \circ \Pi_t^a \circ S_{-\sigma}(Y^+) - \tilde{E}b(Y^+(-\sigma))\| < \frac{\varepsilon}{2}.$

Let $t_7 = \max(t_5,t',t'',t_3)$. Then substitution of (6.8.1-17) and (6.8.1-18) into (6.8.1-14) leads to the desired result.

It remains to show (6.8.1-10). Let t''' be such that for all $t \geq t'''$, $\tau(t) < \delta/2$. For all $t \geq t'''$ let $\tilde{Y}_t^+ := S_{-\tau(t)}(Y_t^+)$. Then, using (i) and (ii) and the Lipschitz condition:

$$\forall r \in [-q+\tfrac{\delta}{2},0]: d^+(\tilde{Y}_t^+(r),Y_t^+(r)) = d^+(Y_t^+(r-\tau(t)),Y_t^+(r)) \leq c.\tau(t).$$

Therefore $\lim_{t\to\infty} \tilde{Y}_t^+|_{[-q+\frac{\delta}{2}]} = Y^+|_{[-q+\frac{\delta}{2},0]}$. Replacing in (6.8.1-9) q by $q-\frac{\delta}{2}$, δ by $\frac{\delta}{2}$ and Y_t^+ by $\tilde{Y}_t^+$, and t_7 by $(t_8:=)\max(t_7,t''')$, one obtains (6.8.1-10).

Q.E.D.

6.8.1-19. Theorem. Let $\{\theta_t^+\}$ be a parameter sequence satisfying property 3 with interpolation function $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ such that $\theta_t^+ = Y^+(s_t)$, for all $t \in \mathbb{Z}$. There exists a set E, $E \subseteq \Omega$, $P(E) = 0$ with the following property. Let $\tilde{S} \in \overset{\circ}{S}$ be such that

(i) $|I(\tilde{S},Y^+)| = 1$; let $\{(i,j)\}$ denote $I(\tilde{S},Y^+)$, and

(ii) $\{S_{r_k}(Y^+)|_{[-q,0]}\}_{k=1}^{\infty}$ converges in $\overset{\circ}{L}{}^+_{ij}[-q,0]$; let $X^+ \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ denote the limit.

Then the following holds

$$\lim_{k\to\infty} f_k = \lim_{k\to\infty} \tilde{E}f_k = \lim_{k\to\infty} \int_{r_k-q+\delta}^{r_k} \tilde{E}b^o(r)dr =$$
$$= \int_o^{q-\delta} \tilde{E}b(X(-\sigma))d\sigma, \quad \forall \omega \in \Omega\backslash E. \tag{6.8.1-20}$$

Proof. Equation (6.8.1-20) contains three equalities that have to be verified The last one will be treated first. From the definition of b^o (cf. (6.6.2-27)) it follows that

$$b^o(r_k-\sigma) = b_t \circ \Pi_t \circ S_{s_t}(Y^+) \tag{6.8.1-21}$$

if $t = t(k)$ is such that $s_t \leq r_k-\sigma < s_{t+1}$ and $t \geq t_o$.
Let $Y_k := S_{r_k}(Y^+)$ and $\tau(k) = r_k - s_{t(k)} - \sigma$ for all k. Then $\lim_{k\to\infty}(Y_k|_{[-q,0]}) = X$ and

$$b^o(r_k-\sigma) = b_t \circ \Pi_t \circ S_{-\sigma-\tau(k)}(Y_k^+), \text{ with } t = t(k). \tag{6.8.1-22}$$

It is not difficult to conclude from theorem (6.8.1-18) (c) that the result presented there holds equally well for a sequence $\{t(k)\}_{k=1}^{\infty}$ instead of all $t \geq t_o$. Applying this one finds that $\{\tilde{E}b^o(r_k-\sigma)\}$ converges uniformly for all $\sigma \in [0,q-\delta]$ to $\tilde{E}b(X^+(-\sigma))$. Therefore

$$\lim_{k\to\infty} \int_o^{q-\delta} \tilde{E}b^o(r_k-\sigma)d\sigma = \int_o^{q-\delta} \tilde{E}b(X^+(-\sigma))d\sigma. \tag{6.8.1-23}$$

(The existence of the integral on the right-hand side follows from the continuity of $\tilde{E}b(X^+(-\sigma))$ as a function of σ). Thus the last equality of

(6.8.1-20) has been shown to hold.
The second equality of (6.8.1-20) follows directly from the definition of f_k and condition (i) on $\tilde{S}$ (cf. (6.6.2-37)). The first equality holds for all $\omega \notin E$ and follows from theorem (6.6.3-16) in combination with theorem (6.7.2-32).

Q.E.D.

From this theorem together with theorem (6.8.1-5) one finds

6.8.1-24. Corollary. Under the conditions of the previous theorem (6.8.1-19),

$$\forall \omega \notin E: \lim_{k\to\infty} f_k = -\frac{1}{2}\int_0^{q-\delta} R(X^+(-\sigma))^{-1}\frac{\partial}{\partial\phi}\tilde{E}g_1(v)\|\varepsilon[\phi_j(X(-\sigma));j)]\|^2 d\sigma.$$

6.8.2. The o.d.e. for the (coupled) algorithm

Consider the 'coupled' algorithm. This means that apart from the equations of the decoupled algorithm (cf. section 6.4) one has equations (6.2.9-11), (6.2.9-12) and the rules for coordinate change that follow (6.2.9-12). For the 'coupled' algorithm one has the following equality.

6.8.2-1. Theorem. Let $\omega \in \Omega$ be fixed. Let $\{\theta_t^+\}_{t=t_o}^{\infty}$ and $\{\hat{\theta}_t^+\}_{t=t_o}^{\infty}$ be the two parameter sequences that are produced by the algorithm (as described in section 6.2.9). Let $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ be an interpolation curve such that $Y^+(s_t) = \theta_t^+$ for all $t \geq t_o$ (such a curve exists according to theorem (6.5-19)). Let $\tilde{S} = (\{r_k\}_1^{\infty}, q, \delta) \in \overset{\circ}{S}$ be such that $|I(\tilde{S}, Y^+)| = 1$, say $I(\tilde{S}, Y^+) = \{(i,j)\}$. Then for all $k \in \mathbb{N}$:

$$(6.8.2\text{-}2)\qquad f_k = \phi_j(Y(r_k)) - \phi_j(Y(r_k - q + \delta)) + \tilde{\tau}(k),$$

and

$$(6.8.2\text{-}3)\qquad \lim_{k\to\infty} \tilde{\tau}(k) = 0.$$

Proof. The proof is given in four steps. In each step a sequence is defined which converges to zero. The sum of those four sequences is $\{\tilde{\tau}(k)\}_{k=1}^{\infty}$ and it then follows that $\lim_{k\to\infty} \tilde{\tau}(k) = 0$.

(i) Let

$$N_k := \{t \mid s_t \in [r_k - q + \delta, r_k]\}, t'(k) := \min N_k \text{ and } t''(k) := \max N_k.$$

Applying proposition (6.6.2-28) to the definition of f_k (6.6.2-37) one finds

$$f_k = \int_{r_k - q + \delta}^{r_k} b^o(\sigma) d\sigma = \sum_{t \in N_k} a_t b_t + \tilde{\tau}_1(k),$$

with

$$\tilde{\tau}_1(k) = [s_{t'(k)} - (r_k - q + \delta)] b_{t'(k)-1} + [r_k - s_{t''(k)+1}] b_{t''(k)}.$$

Now $0 < s_{t'(k)} - (r_k - q + \delta) < a_{t'(k)-1}$ and $0 < r_k - s_{t''(k)+1} < a_{t''(k)+1}$.
Because $a_k \to 0$ for $k \to \infty$ and because $\{b_t\}$ is bounded by a data-independent constant (this follows from corollary (6.4-55)), $\lim_{k \to \infty} \tilde{\tau}_1(k) = 0$.

(ii) Because in the 'coupled' algorithm, $\{\hat{\theta}_t\}$ is well-defined one has (compare (6.6.2-1)) for all $t \in \cup N_k : a_t b_t = \phi(\hat{\theta}_{t+1}, j) - \phi(\hat{\theta}_t, j)$ if $\lambda_t = 1$.
Now $\lambda_t \neq 1$ occurs only if in the next step a coordinate change takes place (cf. (6.2.9-1)ff). Therefore $\lambda_t = 1$ if $t \in N_k \setminus \{t''(k)\}$.
So

$$\sum_{t \in N_k} a_t b_t = \sum_{t=t'(k)}^{t''(k)-1} \{\phi(\hat{\theta}_{t+1}, j) - \phi(\hat{\theta}_t, j)\} + a_{t''(k)} b_{t''(k)} =$$

$$= \phi(\hat{\theta}_{t''(k)}, j) - \phi(\hat{\theta}_{t'(k)}, j) + \tilde{\tau}_2(k),$$

where $\tilde{\tau}_2(k) = a_{t''(k)} b_{t''(k)}$. Clearly $\lim_{k \to \infty} \tilde{\tau}_2(k) = 0$, according to similar arguments as given in (i) above.

(iii) Let (for k sufficiently large such that $\theta_{t''(k)} \in C_j, \theta_{t'(k)} \in C_j$)

$$\tilde{\tau}_3(k) = \{\phi(\hat{\theta}_{t''(k)}, j) - \phi(\theta_{t''(k)}, j)\} - \{\phi(\hat{\theta}_{t'(k)}, j) - \phi(\theta_{t'(k)}, j)\}.$$

From the coupling equation (6.2.9-12) (if also (6.2.9-4) ff), the fact that $\lim_{t \to \infty} \delta_t = 0$ and the equivalence of the inner metric with the local coordinate metric (cf. (6.3.4-23)), it then follows that $\lim_{k \to \infty} \tilde{\tau}_3(k) = 0$.

(iv) Let

$$\tilde{\tau}_4(k) = \{\phi(\theta_{t''(k)},j) - \phi(Y(r_k))\} - \{\phi(\theta_{t'(k)},j) - \phi(Y(r_k-q+\delta))\}.$$

Because Y(r) is an interpolation curve,

$$Y(s_{t''(k)}) = \theta_{t''(k)} \text{ and } Y(s_{t'(k)}) = \theta_{t'(k)}.$$

Because $r_k - s_{t''(k)} \to 0$, $s_{t'(k)} - (r_k - q + \delta) \to 0$, and because Y satisfies a Lipschitz condition (and again using the equivalence of inner metric and local-coordinates-metric), one has $\lim_{k\to\infty} \tilde{\tau}_4(k) = 0$. Now take

$\tilde{\tau}(k) = \tilde{\tau}_1(k) + \tilde{\tau}_2(k) + \tilde{\tau}_3(k) + \tilde{\tau}_4(k)$ and the theorem follows.

Q.E.D.

The combination of this theorem with corollary (6.8.1-24) leads to the differential equation associated with the algorithm.

6.8.2-4. Theorem. There exists a set E, P(E) = 0, with the following property. Let $\omega \in \Omega\backslash E$. Let $\{\theta_t^+\}_{t_o}^{\infty}$ and $\{\hat{\theta}_t^+\}_{t_o}^{\infty}$ be the two parameter sequences that are produced by the algorithm (for this ω). Let $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ be an interpolation curve such that $Y^+(s_t) = \theta_t^+$ for all $t \geq t_o$. Let for all $\delta \in (0,q)$, $\tilde{S} = (\{r_k\},q,\delta) \in \overset{\circ}{S}$ be such that

(i) $|I(\tilde{S},Y^+)| = 1$; let $\{(i,j)\}$ denote $I(\tilde{S},Y^+)$ and

(ii) $\{S_{r_k}(Y^+)|_{[-q,0]}\}_{k=1}^{\infty}$ converges in $\overset{\circ}{L}{}^+_{ij}[-q,0]$; let $X^+ \in \overset{\circ}{L}{}^+_{ij}[-q,0]$ denote the limit.

Then the following holds

$$\forall \delta \in (0,q): \phi_j(X(0)) - \phi_j(X(-q+\delta)) =$$

(a)
$$= -\frac{1}{2} \int_{-q+\delta}^{0} R(X^+(\sigma))^{-1} \frac{\partial}{\partial \phi} \tilde{E}g_1(v) \|\varepsilon[\phi_j(X(\sigma));j]\|^2 d\sigma.$$

(b) For all $r \in (-q,0)$:- $\frac{\partial}{\partial r}\phi_j(X(r)) = \frac{1}{2}R(X^+(r))^{-1} \frac{\partial}{\partial \phi} \tilde{E}g_1(v) \|\varepsilon[\phi_j(X(r));j]\|^2$.

(c) For all $r \in (-q,0)$

$$(6.8.2\text{-}5) \qquad \dot{X}(r) = - \nabla_R V_g(X(r)),$$

where $\nabla_R V_g$ denotes the Riemannian gradient of the function V_g (cf. section 6.2.6 and (6.8.1-7)).

Proof. First remark that if δ' is such that $(\{r_k\},q,\delta')$ satisfies (i) and (ii) then the same holds for all $(\{r_k\},q,\delta)$ with $\delta \in (0,q)$.

(a) Let $\delta \in (0,q)$ be arbitrary. Application of theorem (6.8.2-1) gives

$$\lim_{k\to\infty} f_k = \lim_{k\to\infty} \phi_j(Y(r_k)) - \phi_j(Y(r_k-q+\delta)) =$$

$$= \phi_j(X(0)) - \phi_j(X(-q+\delta)).$$

Now corollary (6.8.1-24) gives the result.

(b) Let $r = -q+\delta$ with q fixed and δ varying over $(0,q)$. Then $r \in (-q,0)$. Substitution of $\delta = q+r$ in (a) and differentiation with respect to r gives

$$- \frac{\partial}{\partial r}\phi_j(X(r)) = R(X^+(r))^{-1} \frac{\partial}{\partial \phi}\tilde{E}g_1(v)\|\varepsilon[\phi_j(X(r));j]\|^2/2.$$

(c) The vector $\frac{\partial}{\partial r}\phi_j(X(r))$ is the expression in local coordinates of the tangent vector $\dot{X}$. As explained in section 6.2.6, the vector

$$R(X(r))^{-1} \frac{\partial}{\partial \phi} \tilde{E}g_1(v)\|\varepsilon[\phi_j(X(r));j]\|^2/2$$

is the expression in local coordinates of the Riemannian gradient vector of the function V_g (defined in (6.8.1-7)). In a coordinate free notation, (b) can be expressed as

$$\forall r \in (-q,0): -\dot{X}(r) = \nabla_R V_g(X(r)),$$

from which (6.8.2-5) follows by multiplication of both sides with -1.

Q.E.D.

6.8.2-6. Remark. Part (c) of this theorem is a basic result about the asymptotic behaviour of the algorithm from which all (other) results about the asymptotic behaviour will be derived.

Part (c) of this theorem will now be generalized in the sense that condition (i) will be dropped, and (ii) will be changed accordingly.

6.8.2-7. Theorem. There exists a set E, P(E) = 0, with the following property. Let $\omega \in \Omega \backslash E$. Let $\{\theta_t\}$ and $\{\hat{\theta}_t\}$ be the two parameter sequences that are produced by the algorithm (for this ω). Let $Y^+ \in \overset{\circ}{L}{}^+_{cP}$ be an interpolation curve such that $Y^+(s_t) = \theta_t^+$ for all $t \geq t_o$.

(As usual let Y denote the projection of Y^+ in $\overset{\circ}{L}_c$). Let $\tilde{S} = (\{r_k\}, q, \delta) \in \overset{\circ}{S}$ be such that $\{S_{r_k}(Y)|_{[-q,0]}\}_{k=1}^{\infty}$ converges in $\overset{\circ}{L}_c[-q,0]$ and let $X \in \overset{\circ}{L}_c[-q,0]$ denote the limit. Then $\forall r \in (-q,0)$:

(6.8.2-8) $\quad \dot{X}(r) = -\nabla_R V(X(r))$.

Proof. Let $X_k^+(r) = Y^+(r_k - r)$ for all $r \in [-q,0]$, then $X_k^+ \in \overset{\circ}{L}{}^+_{cP}[-q,0]$ and $X_k \in \overset{\circ}{L}_c[-q,0]$. Then $\lim_{k\to\infty} X_k =: X \in \overset{\circ}{L}_c[-q,0]$. The following three remarks, labeled (A), (B) and (C) will be useful.

(A) For each subsequence $\{k(\ell)\}_{\ell=1}^{\infty}$ of $\{k\}_{k=1}^{\infty}$ one has $\lim_{\ell\to\infty} X_{k(\ell)} = X$.

A subsequence of $\{k(\ell)\}_{\ell=1}^{\infty}$ is of course also a subsequence of $\{k\}_{k=1}^{\infty}$. In order not to complicate the notation further, all such subsequences will also be denoted by $\{k(\ell)\}_{\ell=1}^{\infty}$.

(B) Because $X \in \overset{\circ}{L}_c[-q,0]$, X is continuous. Also V_g is continuous (it is even differentiable). Therefore it suffices to show that the differential equation (6.8.2-8) holds on $(-q,0)\backslash F$, F some finite set. This can be shown by considering the corresponding integral equation in local coordinates. To make this clear consider an o.d.e. in $\mathbb{R}^n$: $\dot{x}(t) = f(x(t))$, $\forall t \in (0,t_1) \cup (t_1,t_2)$.

If x is continuous on $[0,t_2]$ and f is continuous then it follows that

$$\forall t \in [0,t_1]: x(t)-x(0) = \int_0^t f(x(\tau))d\tau, \text{ and}$$

$$\forall t \in [t_1,t_2]: x(t) - x(t_1) = \int_{t_1}^t f(x(\tau))d\tau.$$

But then clearly

$$\forall t \in [0,t_2]: x(t) - x(0) = \int_0^t f(x(\tau))d\tau.$$

The right-hand side is differentiable, so the left-hand side is differentiable as well and

$$\forall t \in (0,t_2): \dot{x}(t) = f(x(t)).$$

(C) For $\varepsilon > 0$ let $N_\varepsilon(F)$ denote the ε-neighbourhood of F, i.e. $N_\varepsilon(F) = \{x \mid d(x,F) < \varepsilon\}$. Clearly, if the differential equation holds on $(-q,0)\backslash N_\varepsilon(F)$ for each $\varepsilon > 0$, then it holds on $(-q,0)\backslash F$. According to (B) this is sufficient.

The proof will now be given in three steps.

(i) As noted before, $X \in \dot{L}_c[-q,0]$, so X is continuous. Consider the finite cover $\{U_p\}$ of M from proposition (6.3.2-22). Using the compactness of $[-q,0]$ it is not difficult to show that there exists a finite partition of $[-q,0]$, namely $-q = -q^{(0)} < -q^{(1)} < \ldots < -q^{(N)} = 0$, such that for each $n = 1,2,\ldots,N$, there is a p, such that $X([-q^{(n-1)},-q^{(n)}]) \subseteq U_p$. According to (B) it is sufficient to show that the differential equation (6.8.2-8) holds on each interval $(q^{(n-1)},q^{(n)})$, $n = 1,2,\ldots,N$. For each n the proof will be the same, therefore the proof will only be given for $(-q^{(N-1)},0)$. To simplify the notation, $(-q,0)$ will be used instead of $(-q^{(N-1)},0)$. One has now (without loss of generality) $X([-q,0]) \subseteq U_p$.

(ii) Now consider $\{X_k([-q,0])\}_{k=1}^{\infty}$. Because $\lim_{k\to\infty} X_k = X$ in $\dot{L}_c[-q,0]$, and $X([-q,0] \subseteq U_p$, with $X([-q,0])$ compact and U_p open, there exists a k_1 such

that for all $k \geq k_1$, $X_k([-q,0]) \subseteq U_p$. According to proposition (6.3.2-22), X_k^+ (which is an element of $\mathring{L}^+_{cP}[-q,0]$) can have at most two coordinate changes occurring on $[-q,0]$. So three cases can be distinguished, namely: (a) no coordinate changes occur, (b) one coordinate change occurs and (c) two coordinate changes occur. In case (a) one has a finite number of possibilities for the coordinate chart index (i,j); because $(i,j) \in \tilde{I} := \{(i,j) | j \in J, i \in I(j)\}$, the number of possibilities is in fact $|\tilde{I}|$. In case (b) one can distinguish again a finite number of possibilities, namely a first coordinate chart index, (i_1,j_1) on an interval $[-q,-q_1(k))$ (for some appropriately chosen $-q_1(k)$) and a second one (i_2,j_2) on the interval $(-q_1(k),0]$. So the number of possibilities is in fact smaller than or equal to $|\tilde{I}|^2$. Similarly in case (c), the number of possibilities is smaller than or equal to $|\tilde{I}|^3$. The total number of possibilities is therefore smaller than or equal to $|\tilde{I}| + |\tilde{I}|^2 + |\tilde{I}|^3 < \infty$. Because the number of possibilities is finite, at least <u>one</u> of them must occur infinitely often. Choose a subsequence $\{k(\ell)\}_{\ell=1}^{\infty}$ such that $X^+_{k(\ell)}$ has the same possibility for its coordinate chart indices for <u>all</u> $\ell \in \mathbb{N}$.

(iii) If (a) is the case, then there is an $(i,j) \in \tilde{I}$ such that $X^+_{k(\ell)} = (X_{k(\ell)},i,j)$ for all $\ell \in \mathbb{N}$. Then the previous theorem can be applied, with $\{r_{k(\ell)}\}_{\ell=1}^{\infty}$ instead of $\{r_k\}_{k=1}^{\infty}$ and it follows that X satisfies the differential equation. If (b) is the case, $-q_1(k(\ell)) \in [-q,0]$ is the number at which a coordinate change takes place in $X^+_{k(\ell)}$. Because $[-q,0]$ is compact, $\{-q_1(k(\ell))\}_{\ell=1}^{\infty}$ has a convergent subsequence. As explained in (A) this subsequence will again be denoted by $\{-q_1(k(\ell))\}_{\ell=1}^{\infty}$. Let its limit be $-q_1$. For each $\varepsilon > 0$ there exists an ℓ_1 such that for all $\ell \geq \ell_1$, $-q_1(k(\ell)) \geq -q_1-\varepsilon$ and therefore for all $\ell \geq \ell_1$, $X^+_{k(\ell)}|_{[-q,-q_1-\varepsilon]}$ has only <u>one</u> coordinate chart index. Therefore the previous theorem can be applied to $X|_{[-q,-q_1-\varepsilon]}$ and so $X|_{(-q,-q_1-\varepsilon)}$ satisfies the differential equation in this case. A similar argument shows that $X|_{(-q_1+\varepsilon,0)}$ satisfies the differential equation in this case (b). Because $\varepsilon > 0$ was chosen arbitrarily, according to (C) the proof is complete for this case (b). If (c) is the case one can use a similar argument as in case (b).

Q.E.D.

From this theorem the following main result of this section can now easily be derived.

6.8.2-9. Theorem. There exists a set $E, P(E) = 0$, with the following property. Let $\omega \in \Omega \backslash E$. Let $\{\theta_t\}$ and $\{\hat{\theta}_t\}$ be the parameter sequences that are produced by the algorithm (for this ω). Let $Y \in \dot{L}_c[0,\infty)$ be an interpolation curve such that $Y(s_t) = \theta_t$ for all $t \geq t_o$. Let $\{r_k\}$ be a monotonically increasing divergent sequence of positive real numbers.

(a) Let $q_1, q_2 \in \mathbb{R}$, $q_1 < q_2$, and let k_1 be such that $r_{k_1} + q_1 \geq 0$. Let

$$X_k := S_{r_k}(Y)\big|_{[q_1,q_2]} \text{ for all } k \geq k_1.$$

If $\{X_k\}$ converges to X in $\dot{L}_c[q_1,q_2]$ then

$$\dot{X}(r) = -\nabla_R V_g(X(r)) \text{ for all } r \in (q_1,q_2).$$

(b) Let $X_k = S_{r_k}(Y)\big|_{[0,\infty)}$ for each $k \in \mathbb{N}$. If $\{X_k\}$ converges to X in $\dot{L}_c[0,\infty)$, then

$$\dot{X}(r) = -\nabla_R V_g(X(r)) \text{ for all } r \in (0,\infty).$$

Proof. (a) This follows directly from the previous theorem by taking $q = q_2 - q_1$, by taking a subsequence $\{r_{k(\ell)}\}_{\ell=1}^{\infty}$ such that $r_{k(\ell+1)} - r_{k(\ell)} > q+\delta$ for each ℓ (such that $\tilde{S} = (\{r_{k(\ell)}\}_{\ell=1}^{\infty}, q, \delta) \in \dot{S}$) and by shifting the curves over a length q_2.

(b) Take $q_1 = 0$. For each $q_2 > 0$, it follows from (a), by restricting the X_k and X to the interval $[0,q]$, that

$$\dot{X}(r) = -\nabla_R V_g(X(r)) \text{ for all } r \in (0,q_2).$$

Because this holds for each $q_2 > 0$, one can conclude that

$$\dot{X}(r) = -\nabla_R V_g(X(r)) \text{ for all } r > 0.$$

Q.E.D.

6.9. The asymptotic behaviour of the algorithm

In this section the results of the previous section will be used to derive convergence results for the algorithm. To start with consider the following definition.

6.9-1. Definition. Let M_o be a metric space with metric d.

(a) Let $\{x_k\}_{k=k_o}^{\infty}$ be a sequence of points in M_o. A point $x \in M_o$ will be called a limit point if for each $\varepsilon > 0$ and for each $k_1 \geq k_o$ there exists a $k \geq k_1$ such that $d(x_k,x) < \varepsilon$.

(b) Let $X: [0,\infty) \to M_o$ be a function. A point $x \in M_o$ will be called a limit point at infinity if for each $\varepsilon > 0$ and for each $r_1 \geq 0$ there exists an $r \geq r_1$ such that $d(X(r),x) < \varepsilon$.

6.9.2. Remarks. (i) In case (a), x is a limit point of $\{x_k\}$ iff there exists a subsequence of $\{x_k\}$ with limit x.

(ii) In case (b), x is a limit point at infinity iff there exists a sequence $\{r_k\}$ with $\lim_{k\to\infty} r_k = \infty$ such that $\lim_{k\to\infty} X(r_k) = x$.

Now let $\{\theta_t\}_{t=t_o}^{\infty}$ and $\{\hat{\theta}_t\}_{t=t_o}^{\infty}$ be the parameter sequences produced by the algorithm (for given ω) and let $Y \in \overset{\circ}{L}_c[0,\infty)$ be an interpolation function with $Y(s_t) = \theta_t$ for all $t \geq t_o$ (such an interpolation function exists, cf. (6.5-19)).

6.9-3. Lemma. (a) The set $\overset{\circ}{X}$ of limit points of $\{\theta_t\}_{t=t_o}^{\infty}$ is equal to the set of limit points of $\{\hat{\theta}_t\}_{t=t_o}^{\infty}$ and equal to the set of limit points at infinity of Y.

(b) The set $\overset{\circ}{V}$ of limit points of $\{V_g(\theta_t)\}_{t=t_o}^{\infty}$ is equal to the set of limit points of $\{V_g(\hat{\theta}_t)\}_{t=t_o}^{\infty}$ and equal to the set of limit points at infinity of $V_g \circ Y$.

Proof. (a) Because of the coupling equation, $d(\theta_t,\hat{\theta}_t) \to 0$ for $t \to \infty$ and therefore any limit point of $\{\theta_t\}$ is a limit point of $\{\hat{\theta}_t\}$ and vice versa. Because $Y(s_t) = \theta_t$ for all $t \geq t_o$, any limit point of $\{\theta_t\}$ is a limit point at infinity of Y. On the other hand if x is a limit point at infinity of Y then there exists a sequence $\{r_k\}$, with $r_k \to \infty$ for $k \to \infty$, such that

$Y(r_k) \to x$ for $k \to \infty$. For each k let $t(k)$ be such that $|s_{t(k)} - r_k| < a_{t(k)}$. Because of the Lipschitz condition, one has $d(Y(r_k), Y(s_{t(k)})) < c.a_{t(k)}$. Because $a_{t(k)} \to 0$ for $k \to \infty$, one has $Y(s_{t(k)}) \to x$ for $k \to \infty$. But $Y(s_{t(k)}) = \theta_{t(k)}$ and it follows that x is a limit point of $\{\theta_t\}$ (cf. remark (6.9-2)).

(b) Because V_g is continuous on M and M is compact, V_g is uniformly continuous. One can now proceed analogously to the proof of (a). The details are left to the reader.

Q.E.D.

Because M, the manifold of parameter points, is compact and V_g continuous, one has in fact

6.9-4. Lemma. (a) $\overset{\circ}{V} = V_g(\overset{\circ}{X})$, (b) $\overset{\circ}{V} \neq \emptyset$ and (c) $\overset{\circ}{V}$ is compact.

Proof. (a) Because V_g is continuous it is clear that $V_g(\overset{\circ}{X}) \subseteq \overset{\circ}{V}$. On the other hand if $v \in \overset{\circ}{V}$ then there exists a sequence $\{t(k)\}_{k=1}^{\infty}$ such that $V_g(\theta_{t(k)}) \to v$. Because $\theta_{t(k)} \in M$ and M is compact, $\{\theta_{t(k)}\}$ has a convergent subsequence, with some limit $\theta \in \overset{\circ}{X}$. Clearly $V_g(\theta) = v$ and so $v \in V_g(\overset{\circ}{X})$. This shows that $\overset{\circ}{V} \subseteq V_g(\overset{\circ}{X})$ and so $\overset{\circ}{V} = V_g(\overset{\circ}{X})$.

(b) Because M is compact, $\overset{\circ}{X} \neq \emptyset$, so $\overset{\circ}{V} \neq \emptyset$.
(c) $\overset{\circ}{X}$ is closed and $\overset{\circ}{X} \subseteq M$, so $\overset{\circ}{X}$ is compact. Therefore $\overset{\circ}{V} = V_g(\overset{\circ}{X})$ is compact.

Q.E.D.

Because $\overset{\circ}{V}$ is also the set of limit points at infinity of the continuous function $V_g \circ Y$ one can show the following.

6.9-5. Lemma. $\overset{\circ}{V}$ is convex.

Proof. It has to be shown that if $v_1, v_2 \in \overset{\circ}{V}$, $v_1 < v_2$ and $v \in \mathbb{R}$ such that $v_1 < v < v_2$, then $v \in \overset{\circ}{V}$. So assume $v_1, v_2 \in \overset{\circ}{V}$, $v_1 < v_2$ and $v_1 < v < v_2$. Then there is a monotonically increasing divergent sequence $\{r_k\}_1^{\infty}, r_k \in [0,\infty)$ for each $k \in \mathbb{N}$, such that $V \circ Y(r_{2k+1}) \to v_1$ and $V \circ Y(r_{2k}) \to v_2$ for $k \to \infty$. Because $v_1 < v < v_2$ it follows that there exists a k_1 such that for all $k \geq k_1$, $V \circ Y(r_{2k+1}) < v$ and $V \circ Y(r_{2k}) > v$. Because $V \circ Y$ is continuous it follows that for each $k \geq k_1$ there exists a number $s_k \in (r_{2k}, r_{2k+1})$ with

$V \circ Y(s_k) = v$. Therefore $v \in \overset{\circ}{V}$.

Q.E.D.

6.9.6. Corollary. $\overset{\circ}{V}$ is a nonempty, compact interval $[v_1, v_2] \subseteq \mathbb{R}$, with $v_1 \leq v_2$.

Similar results can be obtained for $\overset{\circ}{X}$.

6.9-7. Theorem. The set $\overset{\circ}{X}$ is nonempty, compact and connected.

Proof. Because M is compact, the sequence $\{\theta_t\}$ must have a limit point, so $\overset{\circ}{X} \neq \emptyset$. The set $\overset{\circ}{X} \subseteq M$ of limit points of $\{\theta_t\}$ is closed and therefore compact. Suppose $\overset{\circ}{X}$ is not connected, to obtain a contradiction. Then there are two disjoint open sets U_1 and U_2 that cover $\overset{\circ}{X}$ in a nontrivial way, i.e. $\overset{\circ}{X} \subseteq U_1 \cup U_2$ and $\overset{\circ}{X} \cap U_1 \neq \emptyset$ and $\overset{\circ}{X} \cap U_2 \neq \emptyset$. Let $W = M \backslash (U_1 \cup U_2)$, then W is compact. Because $\overset{\circ}{X} \cap U_1$ and $\overset{\circ}{X} \cap U_2$ are nonempty, for each $\bar{r} > 0$, the interpolation curve $Y(\bar{r}+r)$, $r \geq 0$, must switch back and forth between U_1 and U_2 (infinitely often). Therefore for each $\bar{r} > 0$ the curve $Y(\bar{r}+r)$, $r \geq 0$, enters W (infinitely often). Because W is compact it follows that Y must have a limit point at infinity in W, i.e. $\overset{\circ}{X} \cap W \neq \emptyset$. This is in contradiction with $\overset{\circ}{X} \subseteq U_1 \cup U_2$. Therefore $\overset{\circ}{X}$ is connected. Q.E.D.

Notice that the differential equation has not been used yet. This will be done next. Let $\{r_k\}_1^\infty$ be a divergent monotonically increasing sequence of positive real numbers. Let

$$X_k := S_{r_k}(Y)\big|_{[0,\infty)}, \text{ i.e. } X_k(r) = Y(r_k+r), \forall r \geq 0, \forall k \in \mathbb{N}.$$

$\{X_k\}$ will be called a sequence of translations of the interpolation curve Y.

6.9-8. Lemma. Let $E \subseteq \Omega$ with $P(E) = 0$ be as in theorem (6.8.2-9) and let $\omega \notin E$. Let $X \in \overset{\circ}{L}_c[0,\infty)$ be the limit of a sequence of translations of the interpolation curve, as just described. Then $V_g \circ X: [0,\infty) \to \mathbb{R}$ is monotonically nonincreasing.

Proof. According to theorem (6.8.2-9) (b) X is a solution of the differential equation $\dot{X} = -\nabla_R V_g(X)$.

Therefore one has

$$(6.9\text{-}9) \qquad \frac{\partial}{\partial r} V_g(X(r)) = \langle \nabla_R V_g(X(r)), \dot{X}(r) \rangle_R = -\langle \dot{X}(r), \dot{X}(r) \rangle_R \leq 0,$$

where $\langle , \rangle_R$ denotes the inner product corresponding to the Riemannian metric in the tangent space at a given point of the manifold.

Q.E.D.

A <u>critical point</u> of V_g is (of course) by definition any parameter point θ such that $\nabla_R V_g(\theta) = 0$. A <u>critical value</u> of V_g is a number $v \in \mathbb{R}$ for which a critical point θ exists such that $v = V_g(\theta)$. (Of course if v is a critical value and $v = V_g(\theta)$ for some θ, one <u>cannot</u> conclude that θ is a critical point!). It will now be shown that if X is a limit curve as before, then $V_g \circ X(r)$ is a <u>critical value</u> of V for each $r \geq 0$.

6.9-10. <u>Theorem</u>. Let E, with P(E) = 0, be as in theorem (6.8.2-9) and let $\omega \notin E$. Let $X \in L_c^{\circ}[0,\infty)$ be the limit of a sequence of translations of the interpolation curve, as described before. Then for each $r \geq 0$, $V_g \circ X(r)$ is a <u>critical value</u> of V_g.

Proof. Distinguish two cases.

(i) $V_g \circ X(r)$, $r \in [0,\infty)$, is constant,
(ii) $V_g \circ X(r)$, $r \in [0,\infty)$ is not constant.

ad(i) Because $V_g \circ X(r)$ is constant, one has

$$0 = \frac{d}{dr} V_g \circ X(r) = \langle \nabla_R V_g(X(r)), \dot{X}(r) \rangle_R =$$

$$= -\langle \dot{X}(r), \dot{X}(r) \rangle_R, \text{ for all } r > 0$$

and therefore $\dot{X}(r) = 0$ for all $r > 0$, and
so $\nabla_R V_g(X(r)) = -\dot{X}(r) = 0$ for all $r > 0$, so X(r) is a critical point for all $r > 0$ and by continuity, X(r) is also a critical point for r = 0. Therefore $V_g \circ X(r)$ is a critical value of V_g for each $r \geq 0$. This proves the theorem in case (i).

ad(ii). Now suppose $V_g \circ X(r)$, $r \in [0,\infty)$ is not constant. From lemma (6.9-8) it is known that $V_g \circ X$ is monotonically nonincreasing. It follows that there must be numbers r_1, r_2 with $0 < r_1 < r_2$, such that $V_g \circ X(r_1) > V_g \circ X(r_2)$. Let $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ be such that

$$V_g \circ X(r_1) - \varepsilon_1 > V_g \circ X(r_2) + \varepsilon_2.$$

Consider the open sets (relatively to M)

$$N_1 := \{\theta \in M \mid V_g(\theta) > V_g \circ X(r_1) - \varepsilon_1\},$$

and

$$N_2 := \{\theta \in M \mid V_g(\theta) < V_g \circ X(r_2) + \varepsilon_2\}.$$

Then

$$X(r_1) \in N_1 \text{ and } X(r_2) \in N_2 \text{ and } N_1 \cap N_2 = \emptyset.$$

Now consider the interpolation curve Y and the corresponding real function $V_g \circ Y$. Because $X(r_1)$ and $X(r_2)$ are limit points at infinity of Y, Y(r) enters and leaves N_1 and N_2 infinitely often. Therefore there exists a monotonically increasing divergent sequence of real numbers $\{q_k\}_{k=1}^{\infty}$ such that

(a) $Y(q_{2k+1}) \in \partial N_1$ and $Y(q_{2k}) \in \partial N_2$ for all $k \in \mathbb{N}$, and so

$V_g \circ Y(q_{2k+1}) = V \circ X(r_1) - \varepsilon_1$ and $V_g \circ Y(q_{2k}) = V_g \circ X(r_2) + \varepsilon_2$ for all $k \in \mathbb{N}$, and

(b) the image of the open interval (q_{2k}, q_{2k+1}) is outside N_1 and N_2, i.e.

$$Y[(q_{2k}, q_{2k+1})] \subseteq N_1^c \cap N_2^c.$$

So at each q_{2k}, Y <u>leaves</u> N_2 and at each q_{2k+1}, Y <u>enters</u> N_1.
Now define

(6.9-11) $$T_k := q_{2k+1} - q_{2k}, \quad \forall k \in \mathbb{N}.$$

Two cases will be distinguished:

(1) $\exists T \in \mathbb{R}$: $\liminf_{k\to\infty} T_k = T$, and

(2) $\lim_{k\to\infty} T_k = \infty$.

ad(1) There exists a subsequence of $\{T_k\}$ that converges to T. It is not difficult to see that without loss of generality it may by assumed that $\lim_{k\to\infty} T_k = T$. Consider the sequence of translation curves $\{Z_k\}_{k=1}^{\infty}$ with

(6.9-12) $\quad \forall k \in \mathbb{N},\ Z_k = Y(q_{2k}+r)\big|_{[-\frac{1}{2},T+\frac{1}{2}]} \in \overset{\circ}{L}_c[-\tfrac{1}{2},T+\tfrac{1}{2}].$

From the compactness of $L_c[-\frac{1}{2},T+\frac{1}{2}]$ (cf. theorem (6.5-6)) it follows that the sequence has a convergent subsequence, with limit Z, say. According to theorem (6.8.2-9)(a), Z(r), with $-\frac{1}{2} < r < T+\frac{1}{2}$, is a solution of the differential equation

$$\dot{Z}(r) = -\nabla_R V_g(Z(r)),\ -\tfrac{1}{2} < r < T+\tfrac{1}{2}.$$

Therefore, according to the analog of lemma (6.9-8), Z is monotonically nonincreasing, so

(6.9-13) $\quad Z(0) \geq Z(T).$

On the other hand, by construction,

$$Z_k(0) = V_g \circ X(r_2) + \varepsilon_2 \text{ for each } k \in \mathbb{N}$$

and

$$Z_k(T_k) = V_g \circ X(r_1) - \varepsilon_1 \text{ for each } k \in \mathbb{N},$$

from which it follows easily (using the Lipschitz condition) that

$$Z(0) = V_g \circ X(r_2) + \varepsilon_2$$

and

$$Z(T) = V_g \circ X(r_1) - \varepsilon_1.$$

Therefore $Z(0) < Z(T)$, which contradicts (6.9-12)! The conclusion is that this case cannot occur!

ad(2) In this case consider the sequence of translation curves $\{Z_k\}_{k=1}^{\infty}$ with

$$Z_k = Y(q_{2k}+r)|_{[0,\infty)} \in \overset{\circ}{L}_c[0,\infty). \tag{6.9-14}$$

According to the Arzela-Ascoli theorem $\{Z_k\}$ has a convergent subsequence, with limit Z, say. Z satisfies the differential equation and according to lemma (6.9-8), $V_g \circ Z$ is monotonically nonincreasing. On the other hand, for each $k \in \mathbb{N}$,

$$V_g \circ Z_k(r) \geq V_g \circ X(r_2) + \varepsilon_2 \quad \forall r \in [0,T_k]$$

and

$$V_g \circ Z_k(0) = V_g \circ X(r_2) + \varepsilon_2.$$

From this it follows easily that

$$V_g \circ Z(0) = V_g \circ X(r_2) + \varepsilon_2$$

and

$$V_g \circ Z(r) \geq V_g \circ X(r_2) + \varepsilon_2, \ \forall r \geq 0.$$

But $V_g \circ Z$ is monotonically nonincreasing and therefore $V_g \circ Z(r)$ is <u>constant</u> for all $r \geq 0$. Similarly as in (i) it follows that $Z(0)$ is a critical point of V_g and $V_g \circ Z(0) = V_g \circ X(r_2) + \varepsilon_2$ is a critical value of V_g. Because ε_2 can be chosen <u>arbitrarily</u> from the open interval $(0, V_g \circ X(r_1) - V_g \circ X(r_2))$ it follows that <u>each</u> number in the open interval $(V_g \circ X(r_2), V_g \circ X(r_1))$ is a critical value of V_g. Now the set of critical values of V_g is compact (it is the image of the set of critical points, which is a closed set in M, because $V_g \in C^{\infty} \subseteq C^1$, and therefore compact; it follows

that the set of critical values is compact). It follows that the closed interval

$$[V_g \circ X(r_2),\ V_g \circ X(r_1)]$$

consists of critical values of V_g.
Now $\{r_1,r_2\}$ is an arbitrary pair for which $r_1 < r_2$ and

$$V_g \circ X(r_2) < V_g \circ X(r_1).$$

It follows that for each $r \geq 0$, $V_g \circ X(r)$is a critical value of V_g.

Q.E.D.

6.9-15. Corollary. Let E be as before and $\omega \notin E$. The set $\overset{\circ}{V}$ consists of critical values of V_g.

Proof. Let $v \in \overset{\circ}{V}$. Then there exists a monotonically increasing divergent sequence $\{r_k\}_{k=1}^{\infty}$ such that $V \circ Y(r_k) \to v$ for $k \to \infty$. Let $X_k(r) = Y(r_k+r)|_{[0,\infty)}$ for each $k \in \mathbb{N}$. Then $\{X_k\}_{k=1}^{\infty}$ has a convergent subsequence with limit $X \in \overset{\circ}{L}_c[0,\infty)$ and $V_g \circ X(0) = \lim_{k\to\infty} V_g \circ X_k(0) = v$. From the previous theorem it follows that v is a critical value.

Q.E.D.

This corollary can be combined with corollary (6.9-6) to show that $\overset{\circ}{V}$ consists of one point only, by using the well-known theorem of Sard.

6.9-16. Theorem. Let $\omega \notin E$. The set $\overset{\circ}{V}$ consists of one critical value of V_g.

Proof. Because $V_g: M \to \mathbb{R}$ is C^{∞}, it follows from the theorem of Sard (cf. [Ch-Ha], p. 54ff) that the complement of the set of critical values of V_g is dense in $\mathbb{R}$. Because $\overset{\circ}{V}$ is a closed interval $[v_1,v_2]$, $v_1 \leq v_2$ (according to corollary (6.9-6)) and $\overset{\circ}{V}$ consists of critical values of V_g, it follows that $v_1 = v_2$ must hold and $\overset{\circ}{V}$ is a one-point-set. So $\overset{\circ}{V}$ consists of one critical value of V_g.

Q.E.D.

6.9-17. Corollary. Let $\omega \notin E$. Let $v \in \mathbb{R}$ be such that $\overset{\circ}{V} = \{v\}$. The sequences

$\{V_g(\theta_t)\}_{t_o}^{\infty}$ and $\{V_g(\hat{\theta}_t)\}_{t_o}^{\infty}$ converge to v and $\lim_{r\to\infty} V_g \circ Y(r) = v$.

Proof. Because V_g is continuous and M is compact, $V_g(M)$ is compact. For each t $V_g(\theta_t) \in V_g(M)$. Because $\{V_g(\theta_t)\}$ has only one limit point v, it follows that $\{V_g(\theta_t)\}$ converges to v. Similarly to the proof of lemma (6.9-3) it can be shown that from $V_g(\theta_t) \to v$ for $t \to \infty$ it follows that $V_g(\hat{\theta}_t) \to v$ for $t \to \infty$ and $V_g \circ Y(r) \to v$ for $r \to \infty$.

Q.E.D.

This corollary constitutes already one important result about the asymptotic behaviour of the algorithm. Next, the set $\overset{\circ}{X}$ will be investigated further.

6.9-18. Theorem. Let $\omega \notin E$, E as before. The set $\overset{\circ}{X}$ consists of critical points of V_g.

Proof. Let $\theta \in \overset{\circ}{X}$. There exists a monotonically increasing divergent sequence $\{r_k\}$ such that $\lim_{k\to\infty} Y(r_k) = \theta$. Let $X_k(r) = Y(r_k+r)|_{[0,\infty)}$, then $\{X_k\}$ contains a convergent subsequence with limit X, and $X(0) = \theta$. It follows from the previous corollary that for each $r \geq 0$, $V_g \circ X(r) = v$. Therefore, for all $r > 0$,

$$0 = \frac{d}{dr} V_g \circ X(r) = \langle \nabla_R V_g(X(r)), \dot{X}(r) \rangle_R =$$

$$= -\langle \dot{X}(r), \dot{X}(r) \rangle_R = -\langle \nabla_R V_g(X(r)), \nabla_R V_g(X(r)) \rangle_R .$$

So $0 = \dot{X}(r) = \nabla_R V_g(X(r))$ for each $r > 0$. It follows that X(r) is constant, equal to $X(0) = \theta$ and $\nabla_R V_g(\theta) = 0$, i.e. θ is a critical value of V_g.

Q.E.D.

6.9-19. Corollary. Let $\omega \notin E$. Let $X \in \overset{\circ}{L}_c[0,\infty)$ be the limit of a sequence of translations of the interpolation curve. Then X is a solution of the equations $0 = \dot{X}(r) = \nabla_R V_g(X(r))$, $r \in (0,\infty)$. This implies that X is constant and equal to a critical point θ of V_g such that $V_g(\theta) = v$.

So the differential equation has now been replaced by a static(critical-point-) equation.The main conclusion can now be stated

6.9-20. Main theorem. There exists a subset $E \subseteq \Omega$, $P(E) = 0$, with the

following property. Let $\omega \in \Omega^{*} E$. Let $\{\theta_t\}_{t=t_o}^{\infty}$ and $\{\hat{\theta}_t\}_{t=t_o}^{\infty}$ be the sequences produced by the algorithm.

(a) The sequences $\{V_g(\theta_t)\}_{t=t_o}^{\infty}$ and $\{V_g(\hat{\theta}_t)\}_{t=t_o}^{\infty}$ converge to a critical value v of V_g.

(b) The sequence $\{\theta_t\}_{t=t_o}^{\infty}$ (and $\{\hat{\theta}_t\}_{t=t_o}^{\infty}$ as well) converges to its set of limit points $\overset{\circ}{X}$. $\overset{\circ}{X}$ is a nonempty, compact, connected subset of the set of critical points corresponding to the critical value v of V_g.

6.9-21. Corollary. Let $\omega \notin E$. If all the critical points of V_g are isolated then $\{\theta_t\}_{t_o}^{\infty}$ (as well as $\{\hat{\theta}_t\}_{t_o}^{\infty}$) converges to a critical point θ of V_g and $\{V_g(\theta_t)\}_{t_o}^{\infty}$ converges to the critical value $v = V_g(\theta)$.

6.9-22. Remarks. (i) Whether or not all critical points of V_g are isolated depends on the choice of M. For a given M it could in principle be investigated whether or not V_g has this property. From Morse theory it is known that generically functions have isolated critical points (cf. [Mi])

(ii) From the properties of the steady state Kalman filter it is known that it is the unique filter that leads to the minimal prediction error covariance matrix. Because it has been assumed here that the true system is an element of M, it follows that the criterion function V has a unique global minimum at the true system. Because for large values of K', V_g is only a slight perturbation of V, one may expect V_g to have the same property for large K'. In that case the point θ_o at which V_g has its global minimum is an isolated critical point of V_g and $V_g(\theta_o)$ is of course a critical value. It follows that if $\{V_g(\theta_t)\}$ converges to $V_g(\theta_o)$ then $\{\theta_t\}$ converges to θ_o.

Further remarks and comments will be given in the next section.

6.10. Some final remarks

6.10.1. On the question of global convergence to the true parameter point

In the previous section it was found that the parameter sequence $\{\theta_t\}$ (as well as $\{\hat{\theta}_t\}$) converges to some connected set of critical points of V_g, while one would like to have convergence to the true parameter point i.e. 'consistency' of the algorithm. With respect to this two remarks can be made.

(i) The true parameter point $\tilde{\theta}$ is the unique global minimum of V (this follows from the properties of the steady state Kalman filter). V_g can be considered as a perturbation of V. The size of this perturbation depends on the choice of K' in the definition of g_1 (cf. (6.2.8-4)), especially in relation to the true innovations covariance matrix $\tilde{\Sigma}$. By choosing K' large enough in relation to $\tilde{\Sigma}$, the point(s) θ_o at which V_g take(s) its global minimum can be put arbitrarily close to $\tilde{\theta}$. This is shown in Appendix 6C. In the further discussion it will be assumed that θ_o is close to $\tilde{\theta}$ and - for linguistic simplicity - that θ_o is unique. (If θ_o is not unique the points where V_g takes its global minimum will all be close to $\tilde{\theta}$ for K' large enough and therefore they will be close together. The discussion remains essentially the same if one excludes this possibility).

(ii) The set of critical points of V_g can be split up in (a) the global minimum point θ_o, (b) other local minima and (c) all other critical points, like saddle points and local maxima.

ad(b). It is not clear whether there will be local minima other than the global minimum. The choice of the manifold M of course plays a role in this question. It is in principle possible to check whether one has (almost) arrived at the true parameter value, by comparing the sample covariances (which can be computed on-line) with the theoretical covariance matrices that correspond with the estimated parameter point. (If it becomes more or less clear that the algorithm converges to a non-global, local minimum - or to any other critical point(s) - one might consider the possibility of restarting the algorithm at another parameter point. To this end a variety of random restart procedures familiar from global optimization theory could be employed). Investigation of $V(\theta) = V(\theta;\tilde{\theta},\tilde{\Sigma})$ as a function on $M \times M \times Pos$ could perhaps reveal more about the occurrence of non-global local minima.

ad(c) In this case the situation is in some sense somewhat more hopeful than for the non-global, local minima. Each critical point θ_c of this kind has the following property. In each open neighbourhood of the critical point there are points with lower value of the criterion function V_g than $V_g(\theta)$.

Especially if the Hessian of V_g at such a point is nonsingular, it might very well be possible to show, that the parameter sequence $\{\theta_t\}$ produced by the algorithm will not converge to θ_c. Results in this direction have been treated by Ljung (cf. [Lj 75]). The underlying idea is that because of the noise in the system the parameter sequence will not 'get stuck' on such a critical point. However, there are technical difficulties in the proof and I have so far not succeeded in resolving them. But even if one could show such a result then still in practice one should expect the algorithm to behave rather poorly in the neighbourhood of such a critical point, because the gradient is close to zero. Perhaps a solution to this problem can be found by taking second order information into account. Compare e.g. [Lich] for a comparable situation in the theory of optimization over a manifold. This requires future research. The point of avoiding convergence to critical points of this kind is especially important in the light of the results of [Del 82], which state that 'in most cases', a smooth function on a manifold of fixed McMillan degree must have more than one critical point. This follows from an investigation of the topological structure of such manifolds. The topological theory does not necessarily imply that there will be non-global local minima. But there are results on the minimal number of saddle points and local maxima. These results should warn us that saddle points and local maxima will indeed play a role in the problem of identification within such manifolds. On the other hand, this discussion shows that the claims made by [De - By] that algorithms like the one treated here cannot be globally convergent are (at least) stated in a confusing manner. The problem of global convergence is still open and not settled by their approach.

6.10.2. Some remarks on applications and possible extensions of the algorithm

(i) It should perhaps be stressed again that the algorithm is well-suited for many constrained identification problems. The manifold M is the set of all parameter points that satisfy the constraints. An advantage of this algorithm over other methods like those described in [Ku-Cl] is that one does not need any projection facility, the constraints are satisfied automatically, by construction. Of course, the requirement that M is compact and is a differentiable manifold without boundaries is rather restrictive. Cf. also (v) below, however.

(ii) Due to the rather tolerant formulation of the coupling 'equation' (6.2.9-6) many algorithms, which look rather different at first sight, fall

into the presented class of algorithms (or only small modifications are needed). Examples are the coordinate-free prediction error algorithm presented in [Hnz 85b], and the (generalized) Gauss-Newton algorithm discussed in [Hnz 85a]. (We hope to give more details about these algorithms and their analysis in the near future).

(iii) If M could be allowed to have a nonempty boundary, this would be an important extension of the algorithm. Two kinds of boundaries have to be distinguished. (a) A boundary in the sense of differentiable manifolds, as treated e.g. in [Boo]. The boundary is then itself a manifold (without boundary!) of dimension d-1 (where d is the dimension of M, as usual) and the coordinate charts are homeomorphic to a relatively open subset of the closed half-space $\{x = (x_1,\ldots,x_d) | x \in \mathbb{R}^d, x_1 \geq 0\}$. In this case the extension of the algorithm appears to be rather straightforward. If the parameter point $\hat{\theta}_t$ is at the boundary then, in terms of the local coordinates, one computes the antigradient (i.e. minus the Riemannian gradient) and if it points into the closed half-space it can be used right away, if it points out of the closed half-space then it has to be projected first on the closed half-space, or, equivalently, on the space $\{x=(x_1,\ldots,x_d) | x_1=0\}$, in terms of local coordinates. The practical problem, however, is to find such coordinates. This is in many cases quite difficult, and future research is needed.

(b) A boundary in the topological sense, if M is embedded in some larger topological space. (Case (a) can be regarded as a special case of this one). In this case one allows boundaries of varying dimension (< d). This type of problem will appear for certain constrained identification problems. It also appears if one considers a manifold $M \subseteq M^{m,a,f}_{m,n,m}$ which has lower McMillan-degree systems in its (topological) boundary. As a simple example one can think of the case $M^{m,a}_{1,1,1}$, which is a double, infinite sheeted Riemann surface (cf. chapter 5) in which the origin (i.e. the zero system) forms the boundary. This is clearly not an example of case (a), because the dimension of the manifold is two and the dimension of the boundary is zero. How to handle such cases requires more research.

(iv) A rather straightforward extension is the one to systems with exogenous inputs. One will have to formulate some kind of persistency-of-excitation condition to obtain identifiability of the relevant parameters. There is apparently no conceptual problem to work out this case along the same lines as is done here for the case without exogenous inputs.

(v) Perhaps it will turn out to be possible to generalize the algorithm of

(iv) to the socalled 'external variable' representation of linear systems. In that case the algorithm has to identify the inputs and the outputs from the given external variables. A change in the set of input variables and the set of output variables would then correspond to some kind of change of coordinates. Whether this can be worked out has to be investigated.

(vi) It will also be interesting to consider other probability distributions for the innovations. In many cases the algorithm and the analysis are expected to work the same. A special case is a 'chopped-off' Gaussian distribution. In that case, there seems to be no need to use g_1 and g_2, which simplifies the algorithm. In practice one should in fact use 'chopped-off' Gaussian distributions in many cases, to remain consistent with the available knowledge about the variables under consideration. This is because in many (all?) cases one knows (perhaps rough) upper-and lower bounds for the variables. As a simple example, take the length of a human. This is often presented as a Gaussian variable. But it is clear that zero is a lower bound, and 10m (or the length of the equator,or the diameter of the known universe!) is an upperbound that is clearly sufficient for all practical purposes.

(vii) Another possible way to extend the algorithm (which is also related to (iii)(a) and (b)) is to a more general geometry than the Riemannian geometry used here. Especially one may define a socalled Finsler metric on the space of systems. This means that on each tangent space one has a norm, but not (necessarily) an inner product. In that case it still seems possible to define a gradient direction, and to formulate an analogous algorithm. Also the extension of the analysis may be tractable.

(viii) Recursive identification is closely related to 'tracking', i.e. on-line identification of (slowly) varying systems. To make the algorithm presented here into a tracking algorithm, one just has to change the requirement $\lim_{t\to\infty} a_t = 0$ into $\lim_{t\to\infty} a_t = a_\infty > 0$ (or perhaps $\limsup_{t\to\infty} a_t > 0$). It will be interesting, but probably quite a bit more difficult, to find the asymptotic properties of such an algorithm.

(ix) An advantage of the algorithm is that it can also be considered as an adaptive filtering algorithm. If the parameter sequence converges to the true parameter point, then the algorithm contains the steady state Kalman filter for the system under consideration.

(x) An extension to the continuous time case appears to be quite feasible. In fact, the differential geometric set-up that is used here, is very well suited

for the continuous time case.

(xi) Finally, of course, the proof of the pudding is in the eating and algorithms like this should prove their value in practical applications. Much remains to be done, to implement this kind of algorithm and to compare it with others. This will also be, I hope, the subject of future work.

Appendix 6A. Computation of $T(\theta)$, $\frac{\partial T(\theta)}{\partial \phi_j^k}$ and $\frac{\partial \phi_i}{\partial \phi_j^T}$.

Suppose one does not know (or does not want to compute) the transformation $\phi^i \circ \phi_j^{-1}$ explicitly, nor $T(\theta)$ as a function of θ. Then for the computation of $T(\theta)$, $\frac{\partial T(\theta)}{\partial \phi_j^k}$, $k = 1,2,\ldots,d$, and $\frac{\partial \phi_i}{\partial \phi_j^T}$ one can apply the following method, which makes use only of the knowledge of $(A(\phi,j),B(\phi,j),C(\phi,j)) \in L^{m,a}_{m,n,m}$, at $\phi = \phi_j(\theta)$, its partial derivatives with respect to the components ϕ^k, $k = 1,2,\ldots,d$, at $\phi = \phi_j(\theta)$, $(A(\phi,i), B(\phi,i),C(\phi,i)) \in L^{m,a}_{m,n,m}$, at $\phi = \phi_i(\theta)$, and its partial derivatives with respect to the components ϕ^k, $k = 1,\ldots,d$ at $\phi = \phi_i(\theta)$. There is no claim that the following method is computationally very efficient. The main reason to include it is to show explicitly that knowledge of the quantities just described suffice for the computations at hand.
The method is as follows:

(i) The Jacobian $\frac{\partial \phi_i}{\partial \phi_j^T}$ can be computed as follows. Consider $\theta,\theta_1,\theta_2 \in C_j \cap C_i$, θ being the parameter point at which the change of coordinates has to take place. Consider the difference system

$$(6A\text{-}1) \qquad \left[\begin{pmatrix} A(\theta_1,j) & 0 \\ 0 & A(\theta_2,i) \end{pmatrix}, \begin{pmatrix} B(\theta_1,j) \\ B(\theta_2,i) \end{pmatrix}, \left(C(\theta_1,j), -C(\theta_2,i)\right)\right]$$

(compare chapter 5, (5.2-6)).
Compute the Riemannian metric tensor Q of this set of models at $\theta_1 = \theta_2 = \theta$, with respect to the tangent vector $(\dot{\phi}_j^T, \dot{\phi}_i^T)^T$ (see chapter 5, (5.2-38) (a)); one obtains Q from the formula

$$(6A\text{-}2) \qquad \|(\dot{\phi}_j^T \quad \dot{\phi}_i^T)\|^2 = (\dot{\phi}_j^T \quad \dot{\phi}_i^T) Q \begin{bmatrix} \phi_j \\ \cdot \\ \phi_i \end{bmatrix}.$$

Partition Q as $Q = [Q_1\ Q_2]$, Q_i: $2n \times n$, $i = 1,2$. Then the Jacobian is given by the formula

$$(6A\text{-}3) \qquad J := \frac{\partial \phi_j}{\partial \phi_i^T} = -(Q_1^T Q_1)^{-1} Q_1^T Q_2.$$

This can be shown by using the fact that the kernel of Q is the set

$$\{\begin{pmatrix} Jy \\ y \end{pmatrix} | y \in \mathbb{R}^d\}.$$

One can also partition Q as

$$(6A\text{-}4) \qquad Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}, \; Q_{ij}: n \times n, \; i,j \in \{1,2\},$$

then

$$(6A\text{-}5) \qquad J = -Q_{11}^{-1} Q_{12}.$$

(ii) For $\theta \in M^{m,a}_{m,n,m'}$, $T(\theta)$ can be calculated as follows: Recall the notation for the reachability matrix from section 4.4. From (6.2.7-4) it follows that (with local coordinates ϕ for θ)

$$(6A\text{-}6) \qquad R[A(\phi_i,i),B(\phi_i,i)] = T(\theta)R[A(\phi_j,j),B(\phi_j,j)].$$

From this it follows that

$$\begin{aligned} T(\theta) &= R[A(\phi_i,i),B(\phi_i,i)]R[A(\phi_j,j),B(\phi_j,j)]^T \times \\ &\quad \{R[A(\phi_j,j),B(\phi_j,j)]R[A(\phi_j,j),B(\phi_j,j)]^T\}^{-1} = \\ &= L_{ij}L_{jj}^{-1}, \end{aligned} \qquad (6A\text{-}7)$$

where L_{ij} denotes the unique solution (for given ϕ_i,ϕ_j) of the Lyapunov equation

$$(6A\text{-}8) \qquad L_{ij} - A(\phi_i,i)L_{ij}A(\phi_j,j)^T = B(\phi_i,i)B(\phi_j,j)^T,$$

for all i and j. (So this also defines L_{jj}).
From the formula for $T(\theta)$ one can compute the derivative of $T(\theta)$ with respect to the local coordinate ϕ_i^k:

$$(6A\text{-}9)\qquad \frac{dT(\theta)}{d\phi_i^k} = \frac{\partial(L_{ij}L_{jj}^{-1})}{\partial\phi_i^k} + \sum_{\ell} \frac{\partial(L_{ij}L_{jj}^{-1})}{\partial\phi_j^\ell}\frac{\partial\phi_j^\ell}{\partial\phi_i^k} =$$

$$= \frac{\partial L_{ij}}{\partial\phi_i^k}\cdot L_{jj}^{-1} + \sum_{\ell}\frac{\partial L_{ij}}{\partial\phi_j^k}L_{jj}^{-1}\frac{\partial\phi_j^\ell}{\partial\phi_i^k} - \sum_{\ell} L_{ij}L_{jj}^{-1}\frac{\partial L_{jj}}{\partial\phi_j^\ell}L_{jj}^{-1}\frac{\partial\phi_j^\ell}{\partial\phi_j^k},$$

and $\frac{\partial L_{ij}}{\partial\phi_i^k}, \frac{\partial L_{ij}}{\partial\phi_j^k}$ (now with $i \neq j$) and $\frac{\partial L_{jj}}{\partial\phi_j^\ell}$ are the unique solutions of

respectively, the Lyapunov equations

$$(6A\text{-}10)\quad \frac{\partial L_{ij}}{\partial\phi_i^k} - A(\phi_i,i)\frac{\partial L_{ij}}{\partial\phi_i^k}A(\phi_j,j)^T = \frac{\partial B(\phi_i,i)}{\partial\phi_i^k}\cdot B(\phi_j,j)^T + \frac{\partial A(\phi_i,i)}{\partial\phi_i^k}L_{ij}A(\phi_j,j)^T,$$

$$(6A\text{-}11)\quad \frac{\partial L_{ij}}{\partial\phi_j^k} - A(\phi_i,i)\frac{\partial L_{ij}}{\partial\phi_j^k}A(\phi_j,j)^T = B(\phi_i,i)\frac{\partial B(\phi_j,j)^T}{\partial\phi_j^k} + A(\phi_i,i)L_{ij}\frac{\partial A(\phi_j,j)^T}{\partial\phi_j^k},$$

and

$$(6A\text{-}12)\quad \frac{\partial L_{jj}}{\partial\phi_j^k} - A(\phi_j,j)\frac{\partial L_{jj}}{\partial\phi_j^k}A(\phi_j,j)^T = \frac{\partial B(\phi_j,j)}{\partial\phi_j^k}B(\phi_j,j)^T + B(\phi_j,j)\frac{\partial B(\phi_j,j)^T}{\partial\phi_j^k} +$$

$$\frac{\partial A(\phi_j,j)}{\partial\phi_j^k}L_{jj}A(\phi_j,j) + A(\phi_j,j)L_{jj}\frac{\partial A(\phi_j,j)}{\partial\phi_j^k}.$$

Appendix 6B . Proof of lemma (6.7.2-12)

The idea of the proof is to fix all the components of all the innovations, except one. It will then be shown that a $c > 0$ and an $\varepsilon_o > 0$ can be found as required in the lemma, independent of the values of the fixed innovation-components. From this the result follows directly.
Let $w_t^{(i)}$ denote the i-th component of the innovations vector w_t.
(cf. (6.7.2-2)). Let $\overset{\circ}{F}$, with $\overset{\circ}{F} \subseteq \overset{\circ}{H}$, be the σ-algebra generated by $w_t^{(2)}, w_t^{(3)}, \ldots, w_t^{(p)}, w_{t-1}^{(1)}, w_{t-1}^{(2)}, \ldots, w_{t-1}^{(p)}, w_{t-2}^{(1)}, \ldots, w_{t-2}^{(p)}, \ldots$, i.e. all components of $w_t, w_{t-1}, w_{t-2}, \ldots$, except the component $w_t^{(1)}$. It will be shown that (with $v > 0$ fixed)

(6B-1) $\exists c > 0, \exists \varepsilon_o > 0, \forall \varepsilon \in (0, \varepsilon_o), \forall (s,t) \in \bar{N}, \forall \omega \in \Omega$: $P[v-\varepsilon \leq v_{ts} \leq v \,|\, \overset{\circ}{F}] < c\varepsilon^{\frac{1}{2}}$.

From this it follows, by taking the expectation, that

(6B-2) $\forall c > 0, \exists \varepsilon_o > 0, \forall \varepsilon \in (0, \varepsilon_o)$: $P(v-\varepsilon \leq v_{ts} \leq v) = EP[v-\varepsilon \leq v_{ts} \leq v \,|\, \overset{\circ}{F}] < c\varepsilon^{\frac{1}{2}}$,

and the lemma follows. So it remains to show (6B-1). Let for any real number x,

(6B-3) $(x)_+ := \max(x, 0)$.

If $t = s$ one has $v_{ts} = 0$, so if ε_o is taken smaller than v, then

(6B-4) $P[v-\varepsilon \leq v_t \leq v \,|\, \overset{\circ}{F}] = 0 < c\varepsilon^{\frac{1}{2}}$.

So it remains to prove (6B-1) for $s > t$ (instead of $s \geq t$).
For $s > t$, consider

(6B-5) $P[v-\varepsilon \leq v_{ts} \leq v \,|\, \overset{\circ}{F}] = P[v-\varepsilon-\lambda_1 v_{t-1,s} \leq \|y_{ts}\| \leq v-\lambda_1 v_{t-1,s} \,|\, \overset{\circ}{F}] =$

$= P[(v-\varepsilon-\lambda_1 v_{t-1,s})_+ \leq \|y_{ts}\| \leq v-\lambda_1 v_{t-1,s} \,|\, \overset{\circ}{F}] \leq$

$\leq P[(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 \leq \|y_{ts}\|^2 \leq (v-\lambda_1 v_{t-1,s})^2 \,|\, \overset{\circ}{F}]$.

From (6.7.2-4) it follows that for $t > s$, y_{ts} can be decomposed as follows

$$y_{ts} = y_* + \begin{bmatrix} w_t^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{6B-6}$$

where y_* is $\overset{\circ}{F}$-measurable, and $w_t^{(1)} \sim n(0,\sigma_{11}^2)$, where σ_{11} is the (1,1)-element of Σ (cf. (6.7.2-2)). It follows that, conditional on $\overset{\circ}{F}$,

$$\|y_{ts}\|^2 = f^2 + x^2, \tag{6B-7}$$

where f is $\overset{\circ}{F}$-measurable and chosen to be <u>nonnegative</u> and $x \sim n(x_o,\sigma_{11}^2)$, with x_o $\overset{\circ}{F}$-measurable ($x = w_t^{(1)} + x_o$ in fact). Substituting this one obtains

$$\begin{aligned} &P[(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 \leq \|y_{ts}\|^2 \leq (v-\lambda_1 v_{t-1,s})^2 | \overset{\circ}{F}] = \\ &= P[(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 - f^2 \leq x^2 \leq (v-\lambda_1 v_{t-1,s})^2 - f^2 | \overset{\circ}{F}] = \\ &= P[\{(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 - f^2\}_+^{\frac{1}{2}} \leq |x| \leq \{(v-\lambda_1 v_{t-1,s})^2 - f^2\}_+^{\frac{1}{2}} | \overset{\circ}{F}]. \end{aligned} \tag{6B-8}$$

Because $x \sim n(x_o,\sigma_{11}^2)$, the probability density function of x exists and this function is smaller than or equal to $(2\pi\sigma_{11}^2)^{-\frac{1}{2}}$ for all x. Hence the probability density function of $|x|$ exists and is smaller than or equal to $2(2\pi\sigma_{11}^2)^{-\frac{1}{2}}$.

$$\begin{aligned} &P[\{(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 - f^2\}_+^{\frac{1}{2}} \leq |x| \leq \{(v-\lambda_1 v_{t-1,s})^2 - f^2\}_+^{\frac{1}{2}} | \overset{\circ}{F}] \leq \\ &2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} [\{(v-\lambda_1 v_{t-1,s})^2 - f^2\}_+^{\frac{1}{2}} - \{(v-\varepsilon-\lambda_1 v_{t-1,s})^2 - f^2\}_+^{\frac{1}{2}}] =: \Delta(\varepsilon). \end{aligned} \tag{6B-9}$$

Three disjoint possibilities can be distinguished.

(a) $(v-\varepsilon-\lambda_1 v_{t-1,s})_+ > f$ (≥ 0),
(b) $(v-\varepsilon-\lambda_1 v_{t-1,s})_+ \leq f$ and $(v-\lambda_1 v_{t-1,s}) > f$,

or

(c) $(v-\lambda_1 v_{t-1,s})_+ \leq f$.

ad(a) $(v-\varepsilon-\lambda_1 v_{t-1,s})_+ > f$ implies $(v-\lambda_1 v_{t-1,s})_+ > f$, and $v > f$. In this case

$$\text{(6B-10)} \qquad \Delta(\varepsilon) = 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \int_{(v-\lambda_1 v_{t-1,s}-\varepsilon)_+}^{(v-\lambda_1 v_{t-1,s})_+} \kappa(\kappa^2-f^2)^{-\frac{1}{2}} d\kappa \leq$$

$$\leq 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \int_f^{f+\varepsilon} \kappa(\kappa^2-f^2)^{-\frac{1}{2}} d\kappa = 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \{(f+\varepsilon)^2-f^2\}^{\frac{1}{2}} =$$

$$= 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} (2f+\varepsilon)^{\frac{1}{2}} \leq 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} (2v+\varepsilon_o)^{\frac{1}{2}}.$$

So in this case:

$$\text{(6B-11)} \qquad \forall \varepsilon \in (0,\varepsilon_o): \Delta(\varepsilon) \leq 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} (2v+\varepsilon_o)^{\frac{1}{2}} . \varepsilon^{\frac{1}{2}}.$$

ad(b) $(v-\lambda_1 v_{t-1,s})_+ > f$ implies $v > f$.
Furthermore, in this case $\{(v-\varepsilon-\lambda_1 v_{t-1,s})_+^2 - f^2\}_+ = 0$, so

$$\Delta(\varepsilon) = 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \{(v-\lambda_1 v_{t-1,s})_+^2 - f^2\}^{\frac{1}{2}} =$$

$$\text{(6B-12)} \qquad = 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \int_f^{(v-\lambda_1 v_{t-1,s})_+} \kappa(\kappa^2-f^2)^{-\frac{1}{2}} d\kappa \leq 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} \int_f^{f+\varepsilon} \kappa(\kappa^2-f^2)^{-\frac{1}{2}} d\kappa,$$

and one can proceed as in (a) and obtain (6B-11) for this case too.
ad(c) $(v-\lambda_1 v_{t-1,s})_+ \leq f$ implies $(v-\varepsilon-\lambda_1 v_{t-1,s})_+ \leq f$ and so $\Delta(\varepsilon) = 0$ in this case, and (6B-11) follows trivially in this case. So (6B-11) holds in all cases. Combining (6B-11) with (6B-9) and (6B-5) one finds

$$\text{(6B-13)} \qquad P\left[v-\varepsilon \leq v_{ts} \leq v \middle| \mathring{F}\right] \leq \Delta(\varepsilon) \leq 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} (2v+\varepsilon_o)^{\frac{1}{2}} \varepsilon^{\frac{1}{2}}.$$

Taking $\varepsilon_o \in (0,v)$ arbitrary and $c = 2(2\pi\sigma_{11}^2)^{-\frac{1}{2}} (2v+\varepsilon_o)^{\frac{1}{2}}$, (6B-1) follows.

Q.E.D.

Appendix 6C. On the relationship between V and V_g

The function V_g depends on K' because g_1 does. For fixed $\tilde{\theta}$ and $\tilde{\Sigma}$, the function V_g converges to V in supnorm over M, as $K' \to \infty$. This will first be shown.

6C-1. Lemma. $\forall \delta > 0, \exists K'_o$ such that $\forall K' \geq K'_o$: $\sup_{\theta \in M} |V_g(\theta) - V(\theta)| < \delta$.

Proof. The proof will consist of three steps.
(i) First it will be shown that the set of functions $\{V_g | K' > 0\} \cup \{V\}$ is equicontinuous. Let $\theta_1 \in M$ be fixed. Consider

$f(\theta_1,\theta_2) = \tilde{E}\{\|\varepsilon(\theta_1)\|^2 - \|\varepsilon(\theta_2)\|^2\}$. This is a continuous function of $\theta_2 \in M$ and it is zero if θ_2 is equal to θ_1. Let $N(K') = \{\omega \in \Omega | v > K'\}$, then, by definition of g_1 one has $\omega \in N(K')$ iff $g_1(v) = 0$. Let $N(\infty) = \emptyset$. Then one has

$$V_g(\theta) = \int_{N(K')^c} \|\varepsilon(\theta)\|^2 d\tilde{P}(\omega) \text{ and } V(\theta) = \int \|\varepsilon(\theta)\|^2 d\tilde{P}(\omega).$$

It follows that for arbitrary $\theta_1, \theta_2 \in M$ and for each $K' > 0$,

$$|V_g(\theta_1) - V_g(\theta_2)| = \left| \int_{N(K')^c} \{\|\varepsilon(\theta_1)\|^2 - \|\varepsilon(\theta_2)\|^2\} d\tilde{P}(\omega) \right| \leq$$

$$\leq \int_{N(K')^c} \left| \|\varepsilon(\theta_1)\|^2 - \|\varepsilon(\theta_2)\|^2 \right| d\tilde{P}(\omega) \leq f(\theta_1,\theta_2).$$

And similarly

$$|V(\theta_1) - V(\theta_2)| \leq f(\theta_1,\theta_2).$$

Because $f(\theta_1,\theta_2)$ is continuous as a function of θ_2 and $f(\theta_1,\theta_2) = 0$, it follows that $\{V_g | K' > 0\} \cup \{V\}$ is an equicontinuous set of functions.
(ii) Next it will be shown that for each fixed θ, $V_g(\theta) \to V(\theta)$ as $K' \to \infty$. Consider

$$V(\theta) - V_g(\theta) = \tfrac{1}{2}\tilde{E}\|\varepsilon(\theta)\|^2 - \tfrac{1}{2}\tilde{E}g_1(v)\|\varepsilon(\theta)\|^2 =$$

$$= \tfrac{1}{2} \int_{N(K')} \|\varepsilon(\theta)\|^2 d\tilde{P}(\omega).$$

Because $K_2' > K_1'$ implies $N(K_2') \subseteq N(K_1')$, because $\tilde{P}(\bigcap_{K'>0} N(K')) = 0$ and because the integral $\int_\Omega \|\varepsilon(\theta)\|^2 d\tilde{P}(\omega)$ is convergent (it is equal to $V(\theta)$), it follows that

$$\lim_{K'\to\infty} \frac{1}{2} \int_{N(K')} \|\varepsilon(\theta)\|^2 d\tilde{P}(\omega) = 0,$$

so

$$\lim_{K'\to\infty} V_g(\theta) = V(\theta).$$

(iii) From (i) and (ii) and the compactness of M it follows that $V_g \to V$ in supnorm. (Suppose this is not so. Then there is a $\delta > 0$ and there are sequences $\{K_\ell'\}_{\ell=1}^\infty$ and $\{x_\ell\}_{\ell=1}^\infty$ such that $\forall \ell \in \mathbb{N}$: $|V(x_\ell) - V_{g(v;K_\ell')}(x_\ell)| \geq \delta$.

Because M is compact, $\{x_\ell\}$ has a convergent subsequence with limit $\bar{x} \in M$. Using the equicontinuity it can then be shown that $\limsup_{K'\to\infty} |V(\bar{x}) - V_g(\bar{x})| \geq \delta$, which contradicts the pointwise convergence shown in (ii)).

Q.E.D.

It can now rather easily be shown that, (due to the fact that V has its unique global minimum at the true parameter point $\tilde{\theta}$) V_g has its global minimum at a point (or at points) arbitrarily close to $\tilde{\theta}$ if K' is large enough (for fixed $\tilde{\theta}$ and $\tilde{\Sigma}$).

6C-2. Theorem. (Let $\tilde{\theta}$ and $\tilde{\Sigma}$ be fixed). Let $\varepsilon > 0$ be given. Then there exists a $K_o' > 0$ with the following property. If $K' > K_o'$ and V_g has its global minimum at $\theta_o \in M$, then $d(\theta_o, \tilde{\theta}) < \varepsilon$.

Proof. V takes on its global minimum value at only one point, namely $\tilde{\theta}$. Let $B_\varepsilon(\tilde{\theta}) = \{\theta \in M | d(\theta, \tilde{\theta}) < \varepsilon\}$. Let $V(\tilde{\theta}) + \delta$ be the global minimum of V on the (compact) set $M \backslash B_\varepsilon(\tilde{\theta})$. Then $\delta > 0$, (because $\tilde{\theta} \notin M \backslash B_\varepsilon(\tilde{\theta})$). Take $K_o' > 0$ such that for all $K' > K_o'$, $\sup_{\theta \in M} |V - V_g| < \delta/2$. This is possible according to the previous lemma. Then for all $\theta \in M \backslash B_\varepsilon(\tilde{\theta})$, $V_g(\theta) > (V(\tilde{\theta}) + \delta) - \delta/2 = V(\tilde{\theta}) + \delta/2$, while $V_g(\tilde{\theta}) < V(\tilde{\theta}) + \frac{\delta}{2}$. Therefore V_g takes on its global minimum in the set $B_\varepsilon(\tilde{\theta})$. Q.E.D.

6C-3. Remark. In choosing K' one has to reckon with the scale of the process. If all outputs are multiplied by a constant scaling factor λ, then $\tilde{\Sigma}$ is multiplied with λ^2 and $\tilde{\theta}$ remains the same (!). To obtain the same asymptotic results, K' should be multiplied by λ.

References

[Ab-M] R.A. Abraham and J.E. Marsden 'Foundations of Mechanics' (2nd ed.), Benjamin/Cummings, Reading, Mass., 1978.

[Aka] H. Akaike 'Markovian representation of stochastic processes by canonical variables', SIAM J. Control, 13(1975), pp. 162-173.

[Ama 82] S.-I. Amari, 'Differential geometry of curved exponential families-curvatures and information loss', Ann. Statist, 10 (1982), pp. 375-385.

[Ama 85] S.-I. Amari, 'Differential-Geometrical Methods in Statistics', Lect.Notes in Statistics 28, Springer Verlag, Berlin, 1985.

[Ama 86] S.-I. Amari 'Geometrical theory on manifolds of linear systems', Report Metr. 86-1, Math. Eng. Section, Dept. Math. Eng. and Instr. Phys., Fac. Eng., University of Tokyo, Bunkyo-Ku, Tokyo, Japan.

[An 73] B.D.O. Anderson, 'Algebraic properties of minimal degree spectral factors', Automatica, 9 (1973), pp. 491-500.

[An-M] B.D.O. Anderson and J.B. Moore, 'Optimal Filtering', Prentice Hall, 1979.

[And] T.W. Anderson, 'The Statistical Analysis of Time Series', Wiley, 1971.

[Ao 83] M. Aoki, 'Notes on Economic Time Series Analysis: System Theoretic Perspectives', L.N.Ec. and Math. Syst. 220, Springer Verlag, Berlin, 1983.

[Arn] V.I. Arnold, 'Mathematical Methods of Classical Mechanics', Springer, New York, 1978.

[Au-M] L. Auslander and R.E. MacKenzie, 'Introduction to Differentiable Manifolds', McGraw Hill 1963; Dover, New York, 1977.

[Bau] H. Bauer, 'Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie', 3. Auflage, de Gruyter, Berlin, 1978.

[Bi-C] R.L. Bishop and R.J. Crittenden, 'Geometry of Manifolds', Academic Press, New York, 1964.

[Bo-J] G.E.P. Box and G.M. Jenkins, 'Time Series Analysis: Forecasting and Control' (revised ed.), Holden-Day, San Francisco, 1976.

[Bo-W] O.H. Bosgra and A.J. van der Weiden, 'Input-output invariants for linear multivariable systems', IEEE Trans. Aut. Contr., AC-25(1980), pp. 20-36.

[Boo] W.M. Boothby, 'An Introduction to Differentiable Manifolds and Riemannian Geometry', Academic Press, New York, 1975.

[Bro 70] R.W. Brockett, 'Finite Dimensional Linear Systems', Wiley, New York, 1970.

[Bro 76] R.W. Brockett, 'Some geometric questions in the theory of linear systems', IEEE Trans A.C., AC-21 (1976), pp. 449-455.

[By-D] C.I. Byrnes and T.E. Duncan, 'On certain topological invariants arising in system theory' in: P.J. Hilton and G.S. Young (eds), 'New Directions, in Applied Mathematics', Springer Verlag, New York, 1980.

[By-H] C.I. Byrnes and N.E. Hurt, 'On the moduli of linear dynamical systems', Adv. in Math. Studies in Analysis 4(1979), pp. 83-122.

[By-L] C.I. Byrnes and A. Lindquist (eds), 'Modelling, Identification and Control', North-Holland, Amsterdam, 1986.

[By-M] C.I. Byrnes and C.F. Martin, 'Algebraic and Geometrical Methods in Linear Systems Theory', Amer. Math. Soc., Providence, R.I., 1980.

[Cai 76] P.E. Caines, 'Prediction error identification methods for stationary stochastic processes', IEEE Trans. Aut. Contr., AC-21 (1976), pp. 500-505.

[Cai-H] P.E. Caines and R. Hermann (eds), 'Geometry and Identification: Proc. of APSM Workshop on System Geometry, System Identification and Parameter Estimation', Math. Sci. Press, Brookline, Mass, 1983.

[Ch-Ha] S.-N. Chow and J.K. Hale, 'Methods of Bifurcation Theory', Springer Verlag, Berlin 1982.

[Che] C.T. Chen, 'Linear System Theory and Design', Holt, Rinehart and Winston, 1984.

[Cla] J.M.C. Clark, 'The consistent selection of local coordinates in linear system identification', Proc. J.A.C.C., Purdue 1976, pp. 576-580.

[Co-Le] E.A. Coddington and N. Levinson, 'Theory of Ordinary Differential Equations', McGraw-Hill, New York, 1955.

[Daw] A.P. Dawid, 'Further comments on some comments on a paper by Bradley Efron,' The Annals of Statistics, 1977, vol. 5, no.6, pp. 1249.

[DDH] M. Deistler, W. Dunsmuir and E.J. Hannan, 'Vector linear time series models: corrections and extensions', Adv. Appl. Prob., 10 (1978), pp. 360-372.

[Dei 85] M. Deistler, 'Multivariate time series and linear dynamic systems', to appear in 'Advances in Statistical Methods and Computation', vol.1.

[Del 82] D.F. Delchamps, 'The Geometry of Spaces of Linear Systems with an Application to the Identification Problem', Div. Appl. Sc., Harvard Univ., Cambridge, Mass., June 1982.

[Del 83] D.F. Delchamps, 'Multivariable ARMAX identification: some geometric questions', paper presented at John Hopkins University.

[Del-B] D.F. Delchamps and C.I. Byrnes, 'Critical point behaviour of objective functions defined on spaces of multivariable systems', Proc. 21st IEEE Conference on Decision and Control, Orlando, 1982.

[Den] M. Denham, 'Canonical forms for the identification of multivariable linear systems', pp. 646-655 in [IEEE 74].

[DH] W. Dunsmuir and E.J. Hannan, 'Vector linear time series models', Adv. Appl. Prob. 8(1976), pp. 339-364.

[Doo] J.L. Doob, 'Stochastic Processes', Wiley, New York, 1953.

[Dug] J. Dugundji, 'Topology', Allyn and Bacon, Boston, 1966.

[Fau] P.L. Faurre, 'Stochastic realization algorithms', in [Meh-La].

[FCG] P. Faurre, M. Clerget et F. Germain, 'Opérateurs Rationnels Positifs', Dunod, Paris, 1979.

[Fo 75] G.D. Forney Jr., 'Minimal bases of rational vector spaces with applications to multivariable linear systems', SIAM J. Control, 13 (1975), pp. 493-520.

[Gab] D. Gabay, 'Minimizing a differentiable function over a differentiable manifold', Journal of Optim. Theory and Appl., 37 (1982), pp. 177-219.

[Gan] F.R. Gantmacher, 'The Theory of Matrices', Vol. I and II, Chelsea, New York, 1959.

[Glo 73] K. Glover, 'Structural Aspects of System Identification', Ph. D. dissertation, Dept. of Elec. Eng., M.I.T., Cambr., Mass, 1973, rep. no. ESL-R-516.

[Glo 75] K. Glover, 'Some geometrical properties of linear systems with implications in identification', Proc. IFAC, Boston, Mass, Aug. 1975.

[Glo 84] K. Glover, 'All optimal Hankel norm approximations of linear multivariable systems and their L^{∞}-error bounds', Int. J. Control, 39(1984), pp. 1115-1193.

[Glo-Wi] K. Glover and J.C. Willems, 'Parametrizations of linear dynamical systems: canonical forms and identifiability' in [IEEE 74], pp. 640-645.

[Go-P] G.C. Goodwin and R.L. Payne, 'Dynamic System Identification: Experiment Design and Data Analysis', Academic Press, New York, 1977.

[Goo 85] G.C. Goodwin, 'Some observations on robust estimation and control', in H.A. Barker and P.C. Young (eds.)' IFAC Identification and System Parameter Estimation 1985, York, U.K., 1985', Pergamon Press, Oxford, pp. 851-859.

[Gr-Sz] U. Grenander and G. Szegö, 'Toeplitz Forms and Their Applications', Univ. of Calif. Press, 1958.

[Gu-Me] N.K. Gupta and R.K. Mehra 'Computational aspects of maximum likelinood estimation and reduction in sensitivity function calculations', pp. 774-783 in [IEEE 74].

[Gu-S] I. Guikhman et A. Skorokhod, 'Introduction à la Théorie des Processus Aléatoires', Editions MIR, Moscou, 1980.

[Gu-S 69] I.I. Gihman, A.V. Skorohod, 'Introduction to the Theory of Random Processes', Saunders, Philadelphia, 1969.

[Gui 75] R.P. Guidorzi, 'Canonical structures in the identification of multivariable systems', Automatica 11(1975), pp. 361-374.

[Gui 81] R.P. Guidorzi, 'Invariants and canonical forms for systems, structural and parametric identification', Automatica 17 (1981), pp. 117-133.

[Hah] W. Hahn, 'Stability of Motion', Springer Verlag, Berlin, 1967.

[Han 70] E.J. Hannan, 'Multiple Time Series', John Wiley, New York, 1970.

[Han 71] E.J. Hannan, 'The identification problem for multiple equation systems with moving average errors', Econometrica 39 (1971), pp. 751-767.

[Han 81] E.J. Hannan, 'System identification', in [Haz-Wi], pp. 221-246.

[Han-D] E.J. Hannan, M. Deistler, 'The Statistical Theory of Linear Systems', John Wiley & Sons, New York, 1988.

[Han-K] E.J. Hannan, L. Kavalieris, 'Multivariate linear time series models', Adv. Appl. Prob. 16 (1984), pp. 492-561.

[Haz 76] M. Hazewinkel, 'Moduli and canonical forms for linear dynamical systems III: the algebraic-geometric case' in 'Geometric Control Theory', C. Martin and R. Hermann (eds), Math. Sci. Press, Brookline, Mass., 1977, pp. 291-336.

[Haz 77] M. Hazewinkel, 'Moduli and canonical forms for linear dynamical systems II: the topological case', Math. Syst. Th. 10 (1977), pp. 363-385.

[Haz 79] M. Hazewinkel, 'On identification and the geometry of the space of linear systems', Proc. Bonn Jan. 1979 Workshop on Stochastic Control and Stochastic Differential Systems, Lect. Notes Control and Inf. Sciences, 16, Springer, pp. 401-415.

[Haz 80a] M. Hazewinkel, 'On families of linear systems; degeneration phenomena', pp. 157-190 in [By-M].

[Haz 80b] M. Hazewinkel, '(Fine) moduli (spaces) for linear systems: what are they and what are they good for? 'in: C.I. Byrnes and C.F. Martin (eds), 'Geometrical Methods for the Theory of Linear Systems', D. Reidel, 1980, pp. 125-193.

[Haz 80c] 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 Tracts in Pure and Applied Physics, vol. 4, 1980, pp. 362-404.

[Haz-K 75] M. Hazewinkel and R.E. Kalman, 'Moduli and canonical forms for linear dynamical systems', report 7504, Econometrisch Instituut, Erasmus Universiteit Rotterdam, 1975.

[Haz-K 76] M. Hazewinkel and R.E. Kalman, 'On invariants, canonical forms and moduli for linear, constant finite dimensional dynamical systems' in: G. Marchesini and S.K. Mitter (eds), Proc., Udine, Lecture Notes in Econ. and Math. Syst. 131, Springer Verlag, Berlin, 1976, pp. 48-60.

[Haz-M] M. Hazewinkel and C. Martin, 'Representations of the symmetric group, the specialization order, systems and Grassman manifolds', L'Enseignement Mathématique, 29 (1983), pp. 53-87.

[Haz-Ri] M. Hazewinkel and A.H.G. Rinnooy Kan (eds), 'Current Developments in the Interface: Economics, Econometrics, Mathematics', Reidel, Dordrecht, 1982.

[Haz-Wi] M. Hazewinkel and J.C. Willems (eds), 'Stochastic Systems: The Mathematics of Filtering and Identification and Applications', Reidel, Dordrecht 1981.

[Hel 82] U. Helmke, 'A cell decomposition for the orbit space of the similarity action', Report Nr. 65, F.S. 'Dynamische Systeme', Univ. Bremen, 1982.

[Hel 83] U. Helmke, 'Zur Topologie des Raumes Linearer Kontrollsysteme', Report Nr. 100, F.S. 'Dynamische Systeme', Univ. Bremen, 1983.

[Hel 84] U. Helmke, 'The topology of the space of reachable linear systems I: the complex case', Report 122, F.S. 'Dynamische Systeme', Univ. Bremen, 1984.

[Hel-M] U. Helmke and W. Manthey, 'On closures of spaces of linear dynamical systems', Rep. Nr. 132, F.S. 'Dynamische Systeme', Univ. Bremen, 1984.

[Hi-S] M.W. Hirsch and S. Smale, 'Differential Equations, Dynamical Systems and Linear Algebra', Academic Press, New York, 1974.

[Hnz 85a] B. Hanzon, 'On a Gauss-Newton identification method that uses overlapping parametrizations', in H.A. Barker and P.C. Young (eds), 'IFAC Identification and System Parameter Estimation 1985, York, U.K., 1985', Pergamon Press, Oxford, pp. 1671-1676.

[Hnz 85b] B. Hanzon, 'On a coordinate free prediction error algorithm for system identification', pp. 595-604 in [By-L].

[Hnz 88a] B. Hanzon, 'Riemannian geometry on families of linear systems, the deterministic case', Report 88-62, Faculty of Mathematics and Informatics, Technical University of Delft, 1988.

[Hnz 88b] B. Hanzon, 'A geometric approach to system identification using model reduction techniques', Proc. 8th IFAC/IFORS Symposium on Identification and System Parameter Estimation, Beijing, China, 1988, Pergamon Books, Oxford.

[Hnz-H] B. Hanzon and M. Hazewinkel, 'On identification of linear systems and the estimation Lie-algebra of the associated nonlinear filtering problem' in: G.A. Bekey and G.N. Saridis (eds), 'Identification and System Parameter Estimation, Sixth IFAC Symposium', McGregor and Werner Inc., Washington D.C., 1982, pp. 63-68.

[Hnz-M] B. Hanzon and S.I. Marcus, 'Riemannian metrics on spaces of stable linear systems, with applications to identification', Proc. 21st IEEE Conf. on Dec. and Contr., Orlando 1982.

[Hu] S.T. Hu, 'Elements of General Topology', Holden Day, San Francisco, 1964.

[Hu-W] W. Hurewicz and H. Wallman, 'Dimension Theory', Princeton Univ. Press, Princeton, N.J. 1941.

[Hus] D. Husemoller, 'Fibre Bundles', Springer Verlag, 1975.

[IEEE 74] IEEE Transactions of Automatic Control, vol. AC-19, No. C(1974), Special Issue on System Identification and Time Series Analysis.

[Jaz] A.H. Jazwinski, 'Stochastic Processes and Filtering Theory', Acad. Press, New York, 1970.

[Ka 80] T. Kailath, 'Linear Systems', Prentice Hall, Inc., Englewood Cliffs, N.J. 1980.

[Kaku] S. Kakutani, 'On the equivalence of infinite product measures', Ann. of Math. 49(1948), pp. 214-224.

[Kal 71a] R.E. Kalman, 'Kronecker invariants and feedback', Conf. Ordinary Diff. Equ., June 1971.

[Kal 71b] R.E. Kalman, 'Global structure of classes of linear dynamical systems', Lectures, NATO Adv. Study Inst. on Geom. and Alg. Meth. for Non-linear Systems, London, Aug-Sept, 1971.

[Kal 79] R.E. Kalman, 'On partial realizations, transfer functions, and canonical forms', Acta Polyt. Scand. Ma., 31(1979), pp. 9-32.

[Kal 80] R.E. Kalman, 'Identifiability and problems of model selection in econometrics', 4th World Congress of the Econometric Society, Aix-en-Provence, France, August 30, 1980.

[Kal 82a] R.E. Kalman, 'System identification from noisy data', in: A.R. Bednarek, L. Cesari (eds), 'Dynamical system II', pp. 135-164.

[Kal 82b] R.E. Kalman, 'Identification from real data', pp. 161-196 in [Haz-Ri].

[Kal 83] R.E. Kalman, 'Identifiability and modeling in econometrics', pp. 97-136 in 'Developments in Statistics', vol.4, Acad. Press, New York, 1983.

[Kaz] D. Kazakos, 'Distance measures and estimation performance bounds for continuous-time data', IEEE Trans. Inform. Th. 28(1982), pp. 946-951.

[Kaz-Pa] D. Kazakos and P. Papantoni-Kazakos, 'Spectral distance measures between Gaussian processes', IEEE Trans. Aut. Contr. AC-25(1980), pp. 950-959.

[Kel] J.L. Kelley, 'General Topology', Van Nostrand, 1955.

[KFA] R.E. Kalman, P.L. Falb and M.A. Arbib, 'Topics in Mathematical Systems Theory', McGraw Hill, New York, 1969.

[Ko-N] S. Kobayashi and K. Nomizu, 'The Foundations of Differentiable Geometry', vols. I and II, Wiley (Interscience), New York, 1963.

[Kr-B 77] R.W. Brockett and P.S. Krishnaprasad, 'Scaling rational functions and linear systems identification', Proc. 1977 Conf. Information Sciences Systems, pp. 501-506.

[Kr-B 80] R.W. Brockett and P.S. Krishnaprasad,' A scaling theory for linear systems', IEEE Trans. Aut. Contr. AC-25 (1980), pp. 197-206.

[Kri 77] P.S. Krishnaprasad, 'Geometry of Minimal Systems and the Identification Problem', Ph.D. Thesis, Harvard University 1977.

[Kri 77a] P.S. Krishnaprasad, 'Geometry of parametric models: some probabilistic questions', Proc. 15 Allerton Conf. (1977), pp. 661-670.

[Kri 78] P.S. Krishnaprasad, 'On the geometry of linear passive systems', Charleston meeting of A.M.S., Nov. 1978.

[Ku-Cl] H.J. Kushner and D.S. Clark, 'Stochastic Approximation Methods for Constrained and Unconstrained Systems', Springer Verlag, 1978.

[Lich] A. Lichnewsky, 'Une méthode de gradient conjugé sur des variétés. Application à certains problèmes de valeurs propres nonlinéaires', Numer. Funct. Anal. and Optimiz., 1(5), 1979, pp. 515-560.

[Lin-Pi] A. Lindquist and G. Picci 'Realization theory for multivariate stationary Gaussian processes', SIAM J. Control and Optimization, 23 (1985).

[Lj 75] L. Ljung, 'Theorems for the asymptotic analysis of recursive stochastic algorithms', Report 7522, Dec. 1975, Dept. Aut. Control, Lund Institute of Technology.

[Lj 77] L. Ljung, 'Analysis of recursive stochastic algorithms', IEEE Trans. Aut. Contr., AC-22 (1977), pp. 551-575.

[Lj 78] L. Ljung, 'On recursive prediction error identification algorithms', Report LiTH-ISY-I-0226, Dept. Elec. Eng., Linköping University, Linköping, Sweden.

[Lj 81] L. Ljung, 'Recursive identification', pp. 247-282 in [Haz-Wi].

[Lj 81a] L. Ljung, 'Analysis of a general recursive prediction error identification algorithm', Automatica, 17 (1981), pp. 89-99.

[Lj-Söd] L. Ljung and T. Söderström, 'Theory and Practice of Recursive Identification', MIT Press, Cambridge, Mass., 1983.

[Ma] G.W. Mackey, 'Mathematical Foundations of Quantum Mechanics', W.A. Benjamin, Inc., Reading, Mass., 1963.

[Mak] W. Mak, "Berekening van Riemannse metrieken met REDUCE", Verslag Econometrisch Onderzoek 2A, Econometrisch Instituut, Erasmus Universiteit Rotterdam, maart 1984.

[Mar] M. Marden, 'The Geometry of the Zeros of a Polynomial in a Complex Variable', Amer. Math. Soc., New York, 1949.

[McD] C.C. MacDuffee, 'The Theory of Matrices', Chelsea, New York, 1946.

[Meh] R.K. Mehra, 'Synthesis of optimal inputs for multiinput-multioutput systems with process noise', pp. 211-250 in [Meh-La].

[Meh-La] R.K. Mehra and D.G. Lainiotis (eds), 'System Identification: Advances and Case Studies', Academic Press, New York, 1976.

[Mi] J.W. Milnor, 'Morse Theory', Princeton University Press, Princeton 1963.

[Mi-Go] R.H. Middleton and G.C. Goodwin, 'Improved finite word length characteristics in digital control using delta operators', IEEE Trans.AC, vol. AC-31 (1986), pp. 1015-1021.

[Mue] G. Mueller, 'On Partial Realizations', Dept. Math. Systems Theory, ETH Zürich, sept. 1979, Diplomarbeit bei Prof. R.E. Kalman.

[Nag] E. Nagel, 'The Structure of Science', Routledge and Kegan Paul Ltd., London.

[Ne-P] R. Nevanlinna and V. Paatero, 'Introduction to Complex Analysis', Addison-Wesley, Reading, Mass., 1969.

[Nij] Th.E. Nijman, 'Missing Observations in Dynamic Macroeconomic Modeling', Dissertation, VU Uitgeverij, Amsterdam, 1985.

[No] K. Nomizu, 'Lie Groups and Differential Geometry', The Mathematical Society of Japan, 1956.

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

[Rin] W. Rinow, 'Die Innere Geometrie der Metrischen Räume', Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 105, Springer-Verlag, Berlin usw., 1961.

[Ris 71] J. Rissanen, 'Recursive identification of linear systems', SIAM J. Control, 9 (1971), pp. 420-430.

[Ro] T.J. Rothenberg, 'Identification in parametric models', Econometrica, 39 (1971), pp. 577-591.

[Rob-M] H. Robbins and S. Monro, 'A stochastic approximation method' Annals of Math. Stat., 22 (1951), pp. 400-407.

[Ros] H.H. Rosenbrock, 'State-space and Multivariable Theory', Wiley, New York, 1970.

[Roz] Y. Rozanov, 'Stationary Random Processes', Holden Day, 1967.

[Sch] A.J. van der Schaft, 'System Theoretic Descriptions of Physical Systems', Proefschrift (dissertation), 1983.

[Sar] J.D. Sargan, 'Identification and lack of identification', Econometrica 51 (1983), pp. 1605-1634.

[Seg] G. Segal, 'The Topology of Spaces of Rational Functions', Acta Math. 143, Sept. 1979, pp. 39-72.

[Spi] M. Spivak, 'A Comprehensive Introduction to Differentiable Geometry', vols. I-V, Publish or Perish, Inc., Boston, Mass., 1970.

[Ste] N.E. Steenrod, 'The Topology of Fibre Bundles', Princeton University Press, Princeton, New Jersey, 1951.

[Tan] A. Tannenbaum, 'Invariance and System Theory: Algebraic and Geometric Aspects', LNM 845, Springer Verlag, Berlin, 1981.

[Tig] H.H. Tigelaar, 'Identification and Informative Sample Size', Mathematical Centre Tracts 147, Amsterdam.

[Wae] B.L. van der Waerden, 'Algebra', I, II, Springer, Berlin, Achte Auflage, 1971.

[Wi 78] J.C. Willems, 'Recursive filtering', Statistica Neerlandia, 32 (1978), pp. 1-39.

[Wi 79] J.C. Willems, 'Diktaat Systeemtheorie 79/80', deel I en II, Rijksuniversiteit Groningen, Groningen, 1979.

[Wi 79a] J.C. Willems, 'System theoretic models for the analysis of physical systems', Richerche di Automatica, 10 (1979), pp. 71-106.

[Wi 83] J.C. Willems, 'Input-output and state-space representations of finite-dimensional linear time-invariant systems', Lin. Alg. and its Appl. 50(1983), pp. 581-608.

[Wi 86] J.C. Willems, 'Modelling, approximation and complexity of linear systems', pp. 277-282 in [By-L].

List of mathematical symbols

Ia. Sets-general notation

		Page
$\subseteq$	inclusion	14
$\|V\|$	cardinality of the set V	
$\in$	element of	
$\notin$	not an element of	
$\in\in$		17
$[x]$	equivalence class	17
$V\backslash W = \{v\in V \mid w\notin W\}$		
$V \triangle W=(V\backslash W)\cup(W\backslash V)$,	the symmetric difference	25
V^c	complement of the set V	
∂V	topological boundary of the set V	
$B(\theta,\delta)$	ball with centre θ and radius δ	
$\emptyset$	the empty set	
$\mathbb{N} = \{1,2,3,\ldots\}$	the set of natural numbers	
$\mathbb{N}_0 = \{0,1,2,\ldots\} = \mathbb{N} \cup \{0\}$		
$\mathbb{Z}$	the set of integers	
$\mathbb{Q}$	the set of rational numbers	
$\mathbb{R}$	the set of real numbers	
$\mathbb{R}_+ = \{x \in \mathbb{R} \mid x > 0\}$		
$\mathbb{C}$	the set of complex numbers	
$D(0,1) = \{x \in \mathbb{C} \mid \|z\| < 1\}$		
$\overline{D(0,1)} = \{z \in \mathbb{C} \mid \|z\| \leq 1\}$		
$D(z_0,r) = \{z \in \mathbb{C} \mid \|z-z_0\| < r\}$		

(Ia. Sets - general notation)

		Page
$C(0,1) = \{z \in \mathbb{C} \mid \|z\| = 1\}$		
$Gl_n(\mathbb{R})$	the group of nonsingular nxn real matrices	
C^k	the set of real functions which are k times continuously differentiable	
$C^0([a,b],M)$	the set of all continuous curves $X: [a,b] \to M$	
C^∞		118
C^ω		118
S^2	the unit sphere in $\mathbb{R}^3$	159

Ib. Special sets used in the text

T	time axis	12
U	input alphabet	12
$\overset{\circ}{U}$	set of input functions	12
Y	output alphabet	12
$\overset{\circ}{Y}$	set of output functions	12
$\overset{\circ}{F}$	set of nonanticipative input-output mappings	12
B	set of initial conditions	13
$\overset{\circ}{U}_{\inf T}$		17

(Ib. Special sets used in the text)

		Page
$\overset{\circ}{X}$		17
$\overset{\circ}{X}_{\inf T}$		17
X	state space	17
X_{t0}		20
$\overset{\circ}{G}$	σ-algebra	22
$\{\overset{\circ}{G}_t\}_{t\in T}$	filtration of σ-algebras	23
$\text{Prob}(\overset{\circ}{Y},\overset{\circ}{G})$		23
$\overset{\circ}{U}_0$		28,140
$\overset{\circ}{U}_a$		32
$\overset{\circ}{U}_b$		33
$\overset{\circ}{U}_c$		33
L_B		45
L_F		45
$M(t)$		55
$\text{arma}(p,q)$		83
$\text{arma}(\overline{p},\overline{q})$		83
D_∞		100
$D_\infty(M)$		101
$\overset{\circ}{M}$		101
$S(M)$		102
$\overset{\circ}{M}/\sim$		103
$\overset{\circ}{M}_f$		103

(Ib. Special sets used in the text)

		Page
D		103
M_ψ°		104
N		107
Θ		109
Θ_i		109
(U,ϕ)	coordinate neighbourhood	118
U°		118
B°	coordinate bundle	120
B	bundle space	120
X	base space	120
Y	fibre	120
G	group	120
$L^m_{m',n,m}$		125
$M^m_{m',n,m}$		125
$L^m_{l,m',n,m}$		125
$M^m_{l,m',n,m}$		125
U^m_α		127
W^m_α		127
V^m_α		128
V'^m_α		128
$M^{m,a}_{m',n,m}$		128
$M^{m,a,f}_{m',n,m}$		128
Pos(m)	manifold of all mxm symmetric positive definite matrices	136

(I.b. Special sets used in the text)

		Page
$\overset{\circ}{M}{}^{m,a}_{m',n,m}$		145
T_Σ	tangent space	148
$M^{m,\ell}_{m',n,m}$; $\overset{\circ}{M}{}^{m,\ell}_{m',n,m}$		149
$L^{m,\Delta}_{m',n,m}$		174
$M^{m,\Delta}_{m',n,m}$		174
$\overset{\circ}{R}$		188
M^+; M^-		194
$L^a_{m',n,m}$		199
X_Σ		203
$L^2(\lambda)$		213
Ω	sample space	213
$\overset{\circ}{F}$	σ-algebra	213
$L^2(\lambda)^+$		214
$B(\lambda)$		214
E		229
L		229
C_j	coordinate chart	235
J	index set of cover	235
C'_j; C''_j		236
$J''(\theta)$; $J'(\theta)$		238
E_i; E_{ij}; E'_{ij}; E''_{ij}		251
$I(j)$		251
$\tilde{E}_{ij}$		252
D''_{ij}; D'_{ij}; D_{ij}		252
N'_{ij}; N_{ij}		252
$N^{\frac{1}{2}}_{ij}$		253
U_k		256

(Ib. Special sets used in the text)

	Page
$E_{\alpha j}$	260
θ^{+}	266
$\mathring{L}_c$; $\mathring{L}_c^{+}$; $\mathring{L}_c[a,b]$;	
$\overline{D}_{ij}^{+}$	282
$\mathring{L}_c^{+}[a,b]$; $\mathring{L}_c^{+}(-\infty,b]$; $\mathring{L}_c^{+}[a,\infty)$	282/3
$\mathring{L}_{cP}$; $\mathring{L}_{cP}[a,b]$ etc.	283
$\mathring{L}^{+}$; $\mathring{L}_{ij}^{+}$	283
Γ	291
Φ	291
$\mathring{H}$	291
N_k	301
$\mathring{S}$	302
$I_\ell(\tilde{S}, \{\theta_t^{+}\}_{t=-\infty}^{T})$	303
$I_\ell(\tilde{S},Y^{+})$	303
$I(\tilde{S},Y^{+})$	303
E	313
$\mathring{N} = \{N_i\}_{i=1}^{\infty}$	327
$N_0 = \overset{\infty}{\underset{i=1}{\cup}} N_i$	328
N_k	345
$\mathring{X}$; $\mathring{V}$	352

IIa. Mappings - general notation

		Page
W^T	set of all mappings $T \to W$	
$f: A \to B,\ a \longmapsto b = f(a)$		
$f\|_C$	the mapping f restricted to $C \subseteq A$	
$f\|_{<t} = f\|_{(-\infty,t)}$ etc.		
Dom(f)	the domain of the mapping f	20
$f \cup g$		20

IIb. Special mappings used in the text

F_b	input-output mapping corresponding to the initial condition b	13
F	system function	13
S_δ		14
Φ	state evolution function	17
r	read-out map	17
β	initial condition mapping	17
π	empirical probability	25
S	shift operator	33
$L = S^{-1}$	backward shift or lag operator	42
τ	a backward shift	84
σ	a forward shift	84
$\tilde{\tau}$	a backward shift	84
$\tilde{\sigma}$	a forward shift	84
T_M		101
Ψ		103
$\overline{\Psi}$		105
ϕ		109
ϕ_i		109

(IIb. Special mappings used in the text)

		Page
p_{θ}^{I+1}		111
$p: B \to X$	projection	120
$\beta: L^{m}_{m,n,p} \to M^{m}_{m,n,p}$		131
$\|\Sigma\|$	norm of the system Σ	141
$\overset{\circ}{A}$		144
g_{Σ}	Riemannian metric tensor	148
g^{c}_{Σ}	Riemannian metric tensor	149
vec		155
δ_{Δ}		157
$\Psi(z)$		158
$\rho(z)$; $\phi(z)$		161
$p_{\Delta}(w)$		163
ρ_{α}; ϕ_{β}		167
$\|A\|_{F}$	Frobenius norm of matrix A	168
x;y		170
P_{Δ}		174
$\overline{P}_{\Delta}$		175
g_{Δ}		176
P_{α}		177
F_{Δ}; $\overline{F}_{\Delta}$; F_{β}; $\overline{F}_{\beta}$		178
$\bar{\phi}_{\beta}$		180
$\bar{F}_{\beta}$		181
G_{β}; $\bar{G}_{\beta}$		182
F_{ϕ}		184
d_{M}		186
Q		193

(IIb. Special mappings used in the text)

		Page
K_γ; K_γ^-		194
ϕ_1; ϕ_2; ϕ_3		208
$d(\mu_1,\mu_2)$	Hellinger distance	212
$\langle,\rangle_{L^2(\lambda)}$		213
ψ		214
d_B		214
g_T^F		215
g^F		216
V	criterion function	225
ϕ_j	(manifold-) coordinate map	235
$\tilde{\phi}_j$	bundle-coordinate map	235
$\phi_j = (\phi_j^1, \phi_j^2,\ldots,\phi_j^d)^T$		239
$x(\theta)$; $x(\phi_j,j)$	smooth local section	243
g_1; g_2		246
d_j		261
$\chi_1(t)$; $\chi_2(t)$		269
$X_t^+(r)$; $Y^+(-r)$		284
Π_t		285
Π_t^a; $\tilde{\Pi_t^a}$		286
ρ_c; ρ_c^+; ρ; ρ^+		287
F_k		291
Φ		291
S_{-r}		301

(IIb. Special mappings used in the text)

		Page
$b^o(r)$		302
$f_k = f_{kw}(\{\theta_t^+\}_{-\infty}^T)$		303
Π^r; Π_a^r		304
$\|\|, \|\|_S$	spectral norm	319
$V_g(\theta)$		338
X_k; X		351
$<,>_R$		355

IIIa. Matrices - general notation

		Page
rk(A) = rank(A),	the rank of matrix A	
$\sigma(A) = spec(A)$,	the spectrum of the (square) matrix A, $\sigma(A) = = \{\lambda \in \mathbb{C} \mid \|\lambda I - A\| = 0\}$	
det(A) = $\|A\|$,	the determinant of A	
ℓker(A) =	the left kernel of an nxn matrix A, ℓker(A) = $\{x^T \in \mathbb{R}^n \mid x^T A = 0\}$	
ℓker(H),	H a Hankel matrix	33
im(A)	image of A = $\{x \mid \exists y$ such that $x=Ay\}$	

IIIb. Special matrices used in the text

		Page
R(A,B)	reachability matrix	31
Q(A,C)	observability matrix	31

(IIIb. Special matrices used in the text)

		Page
H_k	Markov matrix	29
A_j		38
B_j		38
$H(z)$	backward transfer matrix	39
$T(s) = H(s^{-1})$	forward transfer matrix	39
$(A(z),B(z))$	backward mfd pair	43
$(A_v(s),B_v(s))$	forward mfd pair	43
$[A(z)\ B(z)]$		44
$[\tilde{A}_v(s)\ \tilde{B}_v(s)]$		46
$[\tilde{A}(z)\ \tilde{B}(z)]$		46
$\Gamma_k = Ey_{t+k}y_t^T$	covariance matrix	63
$\overset{\circ}{H}(H,k_0)$		68
$\overset{\circ}{T}(H,k_0)$		68
$\overset{\circ}{H}(H,k,l,m)$		87
$R_j(A,B)$		125
$Q_j(A,C)$		125
$R(A,B)_\alpha$		126
$(A,B,C,\overline{\Omega})$		135
$L(K)$		141
$M(K)$		142
$L_{A,A(s)}(K)$, etc.		150
$M_{A,A(r)}(K)$, etc.		153
$L_1 := \Delta L$		177
L_c		182
L_0		182
$R = R(\eta)$		198
$G = G(A,B,C,D,\overline{\Omega})$		206
$J = J(A,B,C,D,\overline{\Omega})$		206
Ψ_t		232
$F(\theta)$; $G(\theta)$; $H(\theta)$; K		232

(IIIb. Special matrices used in the text)

	Page
$(A(\phi,j), B(\phi,j), C(\phi,j))$	240
$F(\phi,j), G(\phi,j), H(\phi,j)$	240
$R(\theta,j)$	241
$S(\theta;i,j)$	244
T_{ij}	260
$F(\Phi_j;i,j), G(\Phi_j;i,j), H(\Phi_j;i,j)$	261

IVa. Other symbols - general notation

E	expectation operator	25
χ_g	indicator function of the set G	25
$\|x\|$ with $x \in \mathbb{R}^n$ (or $\mathbb{C}^n$),	the Euclidean (resp. Hermitean norm)	
$\|,\|_{\ell^2}$	the ℓ^2-norm	204
x^T with $x \in \mathbb{R}^n$	the transposed of x	
x^* with $x \in \mathbb{C}^n$	the transposed conjugate of x	
$E(y_{t+1}\|\sigma(y^+))$	the conditional expectation	73
L	the lag operator	
deg p(z), with p(z) a polynomial (matrix)		82
$a \vee b := \max(a,b)$, $a,b \in \mathbb{R}$		87
$a \wedge b := \min(a,b)$, $a,b \in \mathbb{R}$		87
$N(\mu,\sigma^2)$	Gaussian (or normal) distribution with mean μ and variance σ^2	112
$<,>$	inner product	148

IVb. Other special symbols used in the text

Σ	system	12
Δ^+		14
Δ		14

(IVb. Other special symbols used in the text)

		Page
$\frac{1}{n} v_n^{\delta}$	relative frequency	25
$y_{t_0+j\|t_0-1}$		28
t_i		34
s_i		34
$\{\alpha^{(i)}\}$		34
$\{\alpha^{(k)}; b^{(k)}\}$		37
r_i	row degrees	46
ρ_i	row degrees	46
p_k		49
J		91
I		91
d_∞		101
N(I+1)		111
α		163
λ		165
μ		165
β		166
r		186
θ		186
$\Sigma(A,B,C;x)$		202
v_t	standardized innovations	216
w_t	innovations	
θ	parameter	225
ε_t	prediction error	230
a_k	weighting constants	231
$\Psi_{it} = \delta\varepsilon_t/\delta\theta$		231
j = j(r)		238
$\xi(t,j)$; Ψ (t,j); e(t,j) etc.		240
$\sigma(\theta)$; $\lambda_M(\theta)$; λ_0; λ_1		245
v_t		245
K'		245

(IVb. Other special symbols used in the text)

	Page
K_t	246
λ_t	247
δ_t'; δ_t	247
ε	260
$\xi(t;\ i,j)$	261
K_e	262
$I(\gamma)$	262
a_t'	266
θ_t^+	266
$\overline{K}(G)$, $\overline{K}(H)$ etc.	267
λ_2	269
$\bar{\bar{v}}_t$	269
$\overline{K}$	279
K^+	280
t_3	280
$\overline{a}_t$; $\overline{a}$	281
$\tilde{K}^+$; K^+	281
r	283
d^+	286
ω	294
$b_t = b_{t\omega} = b_{t\omega}(\alpha)$	294
s_t	301
$\tilde{S} = (S,q,\delta)$	302
$\tilde{P}$	306
$\tilde{E}$	307
v_{ts}	331
$(y,v,g_1(v),\xi(\theta^+),\varepsilon(\theta),$ $\Psi(\theta,j),\ h(\theta,j),\ b(\theta,j))$	337
$\tilde{\tau}(k)$	344

Author index

Abraham, R.A., 148,224,241, 376
Akaike, H., 376
Amari, S.-I., 139,212,376
Anderson, B.D.O., 66,72,73, 136,137,205,230,376
Anderson, T.W., 376
Aoki, M., 376
Arbib, M.A., 76,79,84,382
Arnold, V.I., 148,376
Auslander, L., 117,148,376
Barker, H.A., 379,381
Bauer, H., 61,135,376
Bekey, G.A., 381
Bishop, R.L., 117,148,376
Boothby,W.M., 117,119,148,226, 228,263,364,377
Bosgra, O.H., 79,376
Box, G.E.P., 376
Brockett, R.W., 132,142,376,382
Byrnes, C.I., 132,133,363,377, 378,380
Caines, P., 377
Chen, C.T., 199,377
Chow, S.-N., 359,377
Clark, D.S., 227,231,246,283, 337,363,383
Clark, J.M.C., 128,224,377
Clerget, M., 72,378
Coddington, E.A., 227,377
Crittenden, R.J., 117,148,376

Dal-Cin, M., 378
Deistler, M., 115,377,378,380
Delchamps, D.F., 131,132,139,363,378
Denham, M., 378
Doob, J.L., 378
Dugundji, J., 282,287,292,378
Duncan, T.E., 132,133,377
Dunsmuir, W., 377,378

Falb, P.L., 76,79,84,382
Faurre, P.L., 72,378
Forney Jr., G.D., 45,46,115,378
Gabay, D., 378
Gantmacher, F.R., 129,258,378
Germain, F., 72,378
Glover, K., 132,133,177,224,378,379
Goodwin, G.C., 157,166,216,219,379
Grenander, U., 379
Guidorzi, R.P., 115,224,379
Guikhman, I. (=Gihman I) 24,107,379
Gupta, N.K., 233,379
Hahn, W., 379
Hale, J.K., 359,377
Hannan, E.J., 115,116,377,378,379,380
Hanzon, B., 139,204,224,240,241,364,381
Hazewinkel, M., 55,79,80,125,126,127, 128,131,132,134,139,379,380,381
Helmke, U., 132,380,381
Hermann, R., 377,379
Hilton, P., 377
Hirsch, M., 381
Hu, S.T., 133,381
Hurewicz, W., 22,381
Hurt, N., 132,377
Husemoller, D., 121,381

Jazwinski, A.H., 381
Jenkins, G.M., 376
Kailath, T., 49,381
Kakutani, S., 213,381
Kalman, R.E., 76,77,79,84,106, 109,127,132,380,382
Kazakos, D., 212,382
Kelly, J.L., 382
Kobayashi, S., 117,119,122,148, 195,204,240,382
Kramer, P., 380
Krishnaprasad, P.S., 132,139,216, 382,383
Kushner, H.J., 227,231,246,283, 337,363,383,
Lainiotis, D.G., 384
Levinson, N., 227,377
Lichnewsky, A. 363,383
Lindquist, A. 377,383
Ljung, L. 224,226,227,229,231, 363,383,384
MacDuffee, C.C., 383
MacKenzie, R.E., 117,148,376
Mackey, G.W., 23,213,383
Mak, W., 221,222,383
Manthey, W., 381
Marcus, S.I., 139,381

Marden, M., 241,245,383
Marsden, J.E., 148,224,376
Martin, C.F., 55,377,379,380
Mehra, R.K., 216,233,379,383, 384
Middleton, R.H., 384
Milnor, J.W., 361,384
Monro, S., 384
Moore, J.B., 66,73,136,137,205,230, 376
Mueller, G., 77,79,80,384
Nagel, E., 384
Neumann, J. von, 213
Nevanlinna, R., 159,160,161,174,180, 196,384
Nijman, Th., 384
Nomizu, K., 117,119,122,148,195,204, 240,382,384
Overbeek, A.J.M., van, 224,384

Paatero, V., 159,160,161,174,180,196,
Papantoni-Kazakos, P., 212,382
Payne, R.L., 216,219,379
Picci, G., 383
Rinnooy Kan, A.H.G., 380
Rinow, W., 138,214,384
Rissanen, J., 224,384
Robbins, H., 384 J., 224,384
Rosenbrock, H.H., 384
Rothenberg, T.J., 219,384
Rozanov, Y., 66,384
Sargan, J.D., 384
Saridis, G.N., 381
Schaft, A.J. van der, 106,384
Segal, G., 132,384
Skorohod, A., 24,107,379
Smale, S., 381
Söderström, T., 226,227,229,231,383
Spivak, M., 117,148,384
Steenrod, N.E., 119,121,122,130,384
Szegö, G., 379
Tannenbaum, A., 385

Tigelaar, H.H., 99,111,112,115,
116,385

Waerden, B.L., van der, 385
Wallman, H., 22,381
Weiden, A.J. van der, 79,376
Willems, J.C., 12,13,17,106,224,
379,380,385
Wittgenstein, L., 100

Young, G.S., 377
Young, P.C., 378,381

Subject Index

A

Abel summable 33

Abelian (group) 166,167,178-181
Absolutely continuous 212,213,214,215
Absolutely convergent 32
Abstract realization 27
Abstract state 31-36,48
Abstract state space 29,37,39
Adaptive filtering 365
A.r.m.a. model representation 7,33,37-41 43,49,57,73,82,86,115
Arma(p,q) model 9,27,76,77,82,83,87-90 114,116
Arma($\overline{p},\overline{q}$) 83-91,94-99,114-116
Arzela-Ascoli theorem 227,284,358
Associated (ordinary) differential equation 228,337,346
Associated principal (fibre) bundle 121, 122,130-132
Asymptotically stable 10,11,63,64,67,68 71,73,93,128-130,135,137,141-149,158, 168,185,198,199,210,218,232,258,259, 329,338
Asymptotically stable matrix 258,319,320
Asymptotically stable norm 258,259
Asymptotically stably invertible 129,130, 136,137,228
Asymptotic behaviour of the algorithm 228,266,337,348,352,360
Asymptotic stability 69,129,204,228,250,258
Auto-isometry 171
Automorphism 180,183

B

Backward mfd pair 43-46,62,63, 82-85,88,90
Backward shift 84
Backward transfer matrix 39,40,43 65,71,84,87-90,93,94.97
Bag 35,48,49
Base space 120-122,131
Block-Hankel matrix 29,33,55,77,78
Borel-Cantelli lemma 60,61,336
Boundary in the sense of manifold theory 226,228,364
Boundary provision 245 87,95
Block-Toeplitz matrix 68
Bundle space 120,121

C

Canonical form 8,115,117,124 127,155,185,191,201,242
Cauchy index 132
Cayley-Hamilton theorem 125
Cesar sum(mable) 33,73
Characteristic function 321
Characteristic polynomial 245
Chebyshev inequality 60,336
Compact-open topology 282,286
Concatenation 12,123
Conformal (mapping) 161,164,180,195
Connection 204
Consecutive measurements 114

Continuous dynamical system 22
Continuous time (system) 12,15,124,139-142, 149,150,153,156,157,167-170,173-176, 185,191-195,198,221,222
Control 105,123
Controllability Grammian 177,180-183
Convergence (analysis) 226-228,234,316,362
Coordinate bundle 119-121,131,132
(Coordinate) chart 8,219,224,234-237,246-250, 255,261,264-266,300,303,309,311,333,339,350
Coordinate neighbourhood 238,239
Coupled algorithm 266,280,283,290,306-308, 344,345
Coupling equation 234,240,247,250,265,268, 345,352,363
Covariance matrix 63,95-97,113,205,225,230, 361,362
Cover(ing) 109,110,118,119,228,235-238,250, 256,260,263,264,293,333,349
Criterion (function) 11,225,361
Critical point 11,227,355,360-363
Critical value 355,358-361
(Cross-) section 122,236,243
Curvature 223

D

Decoupled algorithm 265,266,268,274,279, 280,294,306,307,333,337,338,341,344
δ_Δ-asymptotically stable 171,174,175
δ_Δ-operator 157,158,162,163
δ-shift invariant
Derivative of a system 145,147
Deterministic (linear) (dynamical) system 6-12,23-27,57,58,62,63,76,80,91,95, 106,124,136,139,206-212,221,223
Difference equation 13
Differentiable manifold 8-11,22,117,118,128,224
Differential equation 13
Differential geometry 8,9,117,204
Discrete-time (system) 10,12,15,27 28,124,139-143,149,156,157,166-177,185,191-193,221-223
Distribution function 332
Dominance order 55
Dynamical system 9,12,13,22,27
Dynamical system with s.i.c. (set of initial conditions) 13

E

Econometrics 6
Effective (group of transformations) 119,120
Empirical probability 25,107,111
Empty model 102,106
Engineering science 6
Entier 276
'Equal in the past' 17,21
Equicontinuity/equicontinuous 4,284,292-294,300,301,304-307, 312,373,374
Equilibrium point (of an o.d.e.) 337
Equivalence 125
Equivalence class 123,125,131,136
Equivalence of metrics 262,345,346
Equivalence relation 102,121-124
Equivalent 121-127,135-140
Equivalent representation 13
Ergodicity 9,26,57,58,62,107,111, 135
Ergodic theorem 24,25
Exception/exceptional event 228, 290,293,313
Experiment 104

Exponential decay (e.d.) 4,228,316,320,323, 327,330,341
Exponentially decaying (e.d.-) sequence (of vectors, matrices) 316-320,327,334
Exponentially decaying (e.d.-) double sequence (of matrices etc.) 316,317, 322-326,332,333
Exponential decay of dependence (e.d.d.) 321-334
Exponential decay with respect to the interval sequence $\overset{\circ}{N}$(e.d.i-$\overset{\circ}{N}$) 328,334
Exponential decay of dependence w.r.t. interval sequence $\overset{\circ}{N}$(e.d.d.i-$\overset{\circ}{N}$) 328, 329,333
Extended state 231

F

Falsifiable 102
Falsification 6,106
Falsifying nullset 107
Families of (stochastic) linear systems 117, 135,138,140,205
Fibre 120,121,123,139
Fibre bundle 10,117-122,130,131
Finite cover/covering 109
Finite dimensional (linear) (system) 12,17, 22,27-31,49,57,76,135,141
Finite identifiability (problem) 7,9,27,76, 77,95,100,111
Finsler metric 365
Fisher information matrix 11,139,212,215-220
Fisher metric 11,139,212,215-219
Fortran 156
Forward mfd pair 43-46,86
Forward shift 84
Forward shift invariant 14,15,20
Forward time invariant 14,15,26
Forward transfer matrix 39,40,43,6
Frequency domain 26
Frobenius norm (of a matrix) 167, 168,318,319,324

G

Gaussian (white noise) 26,57,59,63 74,96,97,107,135,211,321,322, 330,365
Geodesic 139,185,186,189,190, 214,240,263
Gradient 138,225,226,230,363,364
Group 120-122,161,166,171,180-183, 194

H

Hankel matrix 7,9,27,37,38,55,56, 68,69,82-94,128,132,200
Hankel-Frobenius norm 143,155,182
Hankel matrix of covariances 9, 68,72
Hankel norm 182
Hellinger affinity 213,214
Hellinger distance/metric 11,139, 212-215
High order coefficient matrix (h.o.c.m.) 47,48
Hilbert space 10,138-142,145-149, 155,189,190
Hurwitz polynomial 129
Hypothesis 308-313,316,336
Hysteresis 258

I

Ideal observation 102,103
Identifiability (problem)/identifiable 6,9,100,103,107,114,115,215,219, 225,364
Identification (algorithm) 6,57,117,218, 226,229,363
Impulse-response matrix 59,76
Indicator function 268
Informative 111,112,116
Informative sample size 6,113-116
Initial conditions 8,12,13,17-28,31,32,44, 45,58,62,273
'Initially equal' 17
Inner geometry 185
Inner metric 10,138,139,177,181,189,190, 214,218,248,250,261,262,345,346
Inner shell 236
Innovations representation 11,67,73,74,97, 135-137,206-211,216-218,228,329,362, 365,370
Input (sequence/components) 22,28,31,58,90, 130,135,140,141,211,273,364,365
Input function 12,140
Input-output (mapping) 12,13,42,58,59,117, 123,125,140,141,148,199,201
Interpolation 247
Interpolation curve 11,227,228,282-284,288, 294,304,312,313,333,338,344-350,354,356
Interpolation function 307,308,314,343,352
(A-)invariant 70
Inverse system 62,63,67,90-92
Isometry 10,139,168,171,174-184,189,193, 194,218

J

Jacobian 148,187,192,244,367
Jordan normal form 259

K

Kullback-Liebler information (distance) 215
Kronecker indices 9,35,49,72,77,89, 115,224

L

Lebesgue measure 314
Left kernel 33,69,86
Left norm (of a matrix) 258
Limit point 352-354,360,361
Limit point at infinity 352-254
Linear (dynamical) system 6,7, 10-13,27,38-47,62,80,84-87, 91,93,96,97,123,130,135,141, 146,156-158,167,168,272,365
Linearity 13,26
Line integral 142,143,173
Lipschitz constant (global) 282-284,289,309,310,324-326, 330,340,342,346,353,357
Local coordinates/parametrization 138,215,217,234,239,241,267, 345-347
Local identifiability/locally identifiable 11,111,139,219,220
Locally Euclidean 117,118,190
Local parameter identifiability 110,111
Log-likelihood 225

Lyapunov equation 58,141,142,149,150,156, 167-171,174-179,202,244,368,369
Lyapunov operator 143

M

Markov matrix 59,69-72,76,77,80,82,87-96, 99,115,136,140,141,146,147,181,196, 201,202,206
Matrix fraction description (mfd-) pair 44-49,72,83,84,88,89,94,115
Maximum likelihood (estimator/problem) 216, 219,225
McMillan degree 48,57,72,76,77,80-83,87,89, 94,96,115,136,139,141,145-147,181,185, 197-202,206,208,224,228,233,363,364
Minimal base form 45-49,57,62,82,83,94
Minimal partial realization 76,81
Minimal realization/representation 31,40,41, 57,58,124,125,130,134,136,158,175,198, 199,211
Minimal set of initial conditions 19
Minimum phase (property) 63,67-71
Missing observations 114
Möbius transformation 157,161
Model identifiability 6,103,111
Model space/set 6-10,57,101,103,107,110,114, 138,139,219,223,225,228
Morse theory 361
Multiple input/multiple output (m.i.m.o.) system 134
Multiple input/single output (m.i.s.o.) system 221
Multivariable (linear dynamical systems) 8,62,79,82,116,139

N

N(I+1) 111
Nice selection 124-131,242
Nonanticipative (input/output mapping) 12,13,20,21,47,58
Norm 142,182,183
Nucleus-double-shell (n.d.s.-) cover 236,237,250-253,256,264
Nucleus-shell (n.s.) cover 251
Null set 309,311

O

Objective function 224,227
Observability 174,211
Observability Grammian 182
Observability index 35,48,49,55-57 66,72,83-89,93,94,115
Observability matrix 31,57,124
Observable 124,135,136,140,175,199 208
Observable variable 100,101
Observation 100,113,299
Observationally equivalent 6,8,24, 102,103
On-line 224,365
Orbit (space) 125,175
Order (of a system) 22,117
Ordinary differential equation (o.d.e.)-method 227,337,344, 348,350,354,357-360
Overlapping parametrizations 111, 224,234
Outer metric 174-177,181,190
Outer shell 236

Output 32,48,130,136,140,141,148,168,211, 228,231,272,329,365
Output alphabet 48
Output function 12

P

P-a.s. (P-almost sure) convergence 228,290
Parallel displacement 11,204
Parameter 100,155,225
Parameter estimate 248
Parameter identifiability 9,108,110,185,191
Parameter sequence 265-268,274,279-285,288, 289,298,299,302,303,307-310,313-315, 337-343,346,362-365
Parameter space 111,140,219,234,241,282,299
Parameter update equation 240,246,267
Parametrization 11,109-111,114-117,138,139, 185-191,197-202,217-224
Parseval relation 142
Partial order 50,51,57
Partial realization 72,76,77,80-82,91,95,96
Partial realization lemma 77,82,90,94,95
PBH-test 174,175
Persistency of excitation 105,364
Poincaré metric 196,197
Pole 217
Prediction error 219,230,233,338,361
Prediction error algorithm 8,11,226-230,338
Principal (fibre) bundle 121,122,132,174,198, 229,235,237,242
Probability distribution/measure 23-25,106, 111-113,135,139,212-215,231,306,337,365
Product bundle 122
Projection 120-123,131
Projection facility 227,229,363
Proper 41,43,66,145
Property 0 266,274,275,279-282,296
Property 1 267,268,279-282,296,315
Property 2 281,282,289
Property 3 289,290,294,308,313-315, 333,343,
ψ-model (or system) identifiable 105,108

Q

Quantum mechanical models/quantum mechanics 23,213
Quotient space 125,126
Quotient topology 126

R

Random initial condition 25
Reachability 174
Reachability index 92-94
Reachability matrix 31,57,58,124, 126,368
Reachable 124,135,136,175,199,208
Read-out map 18,19,21,39
Real analytic manifold 118,128-131
Realization 23,24
Realization theory 76
Recursive identification 8,11,138, 224,226,229,365
Recursive prediction error (r.p.e.) algorithm 233,234
'Reduce' 221,223

Relative topology 282
Riemannian geometry 117,138,139,185,188, 219,263,365
Riemannian gradient (algorithm) 195,224,226 234,241,338,347,364
Riemannian metric (tensor) 7,10,138-140,145, 148,149,155,156,174,176,182,185-198, 202-212,215,219-224,229,241,250,262, 355,367
Riemann sphere 157,160-165,170,173-175
Riemann surface 188,189,193-197,223,364
Right norm (of a matrix) 258
Routh-Hurwitz stability criterion 159

S

Sard, theorem of- 359
Scalar matrix 195
Second order n.d.s. cover 253,260,264
Section, see: cross-section
Sequence of intervals 302,327
Sequence of translations of a curve 354-360
Shift-invariant/invariance 13-15,25,289
σ-algebra 22,25,58,73,107,111,212,213,370
Simply connected 132
Single input/multiple output (s.i.m.o.) system 221
Single input/single output (s.i.s.o.) system 10,81,82,134,139,182,185,199,221
Singular value 258
Space(s) of linear systems 7,117,132
Specialization order 9,49,55,56,89
Spectral density matrix 63-67,96,99,210
Spectral factorization (theorem) 9,63,66,67, 72
Spectral norm 258-261,264,267,270-273, 318,319,333
Spectrum (of a matrix) 32,71,73,128,158,245
Stability region 10,142,157,162, 163,167,172
Stable (deterministic) system 145, 166,258
State (vector) 10,16,20,27,29,36, 58,123,124,140,146,203,204,265
State (vector) bundle 10,11,124, 126,130-132,139,174,202,203, 229,234,235,240,242
State space model/representation 7,16-23,27,28,36,39,40,57,76, 80,89,95
Static system 15,16
Stationary (stochastic process) 24,25,57,62,64,67,71,73,96,97, 100,106,107,111,135
Steady state distribution/random variables 337,338
Steady state Kalman filter 73,137, 211,230,361,362,365
Stochastically independent 321-332
Stochastic approximation (method) 231
Stochastic initial conditions 23
Stochastic inputs 23
Stochastic (linear dynamical) system/model 6-9,12,22-24,57 58,62,63,68,73,95,97,101,106, 111,117,135-139,206-209,212, 215-221,224-228,234
Stochastic realization 24-26
Stochastic state space model 24,57 115
Strictly proper 41,145
Strong stochastic realization 24
Structural index 224
Successor index 126

Supnorm 373,374
Symbolic calculation/computation 155,156,221
System function 13,20
System identifiability 9,100,103-110,114,219
System identification 89,101,110,138,227
System theory 16
System-with-state 122-124,130,202

T

Tangent bundle 229,240
Tangent space 10,138,148,155,198,215
Testable model 101-104
Testable stochastic model 107
Test statistic 106,111
Time axis 12-14
Time invariance/invariant 10-15,26-31,84, 87,93,122,338
Time varying linear system 258
Toeplitz matrix 98
Topological boundary 89
Topological group 119
Topologically equivalent 283
Topological manifold 117,118,133
Topological space 21,32,117
Topology 7,9,89,110,117,125,134,236,251, 282,339
Topology of pointwise convergence 56,286
Transfer function 27,67,132,136
Transfer matrix 39-44,63,142,167,183,184 216
Trivial bundle 122
Trivial model 102,106

U

Unbiased estimator 112,113
Uniform continuity 299,353
Unique partial realization (problem) 76,77,80-82
Unstable (system) 218,258
Update equations (of the algorithm) 247,249

V

Vector bundle 121,122,126,131,137
Vector space 22,32

W

Weak stochastic realization 24
White noise 63,97,225
'With probability one'/w.p.1, see: P-a.s.

Y

Young diagram 50

Z

Zero 217,218,224

Addenda and Errata

Page 14, definition 2.1-8: If $t - \delta \notin T$, then define formally $\omega(t - \delta) := \omega_0$, where ω_0 is some fixed value in U (e.g. $\omega_0 = 0$ if $0 \in U$).

Page 34: A concise definition of p_k, k=1,2,...,m, is as follows: Let Dep(i;H) be the property that row i of the Hankel matrix H depends on the previous rows 1,2,...,i-1. Then

$$p_k := -1 + \min\{p \in \mathbb{N}_0 \mid \text{Dep}(pm + k; H)\}.$$

Alternatively, let Ind(i,H) denote the property that row i of the Hankel matrix is independent of the previous rows 1,2,...,i-1. Then

$$p_k = \max\{p \in \mathbb{N}_0 \mid \text{Ind}(pm + k; h)\}.$$

Page 66/67, theorem 2.4.2-17: The unicity is (of course) within the class of square, real, proper rational (transfer) matrices.

Page 127: Lemma 4.4-5 follows in fact directly from the existence of the controllability (Kronecker) indices. They can be analysed in a completely dual fashion to the analysis of the observability indices, as given in 2.3.3.

Page 145: The metric (5.2-25) is a metric on the set of stable systems (as opposed to the (sub)set of asymptotically stable systems) and not more than that:

(i) If A is stable then $\{A^k\}_{k=0}^{\infty}$ is bounded and therefore $\{CA^{k-1}BB^*(A^*)^{k-1}C^*\}_{k=1}^{\infty}$ is bounded and so

$$\operatorname{tr} \sum_{k=1}^{\infty} \frac{CA^{k-1}BB^*(A^*)^{k-1}C^*}{k^2} \text{ is bounded and well-defined.}$$

(ii) If A is unstable (and (A,B,C) minimal of course) then

(a) $\exists\, \lambda \in \sigma(A)$ with $|\lambda| > 1$ and there is an eigenvector $x \in \mathbb{C}^n \setminus \{0\}$ with $Ax = \lambda x$,

and/or

(b) $\exists\ \lambda \in \sigma(A)$ with $|\lambda| = 1$, there is a corresponding generalized eigenvector $x \in \mathbb{C}^n \setminus \{0\}$ and there is an eigenvector $y \in \mathbb{C}^n \setminus \{0\}$ such that

$$Ax = \lambda x + y$$

In case (a) $\frac{A^k x}{k} = \frac{\lambda^k}{k} x$ and $\left|\frac{\lambda^k}{k}\right| \to \infty$ if $k \to \infty$.

In case (b) $\frac{A^k x}{k} = \frac{\lambda^k}{k} x + \frac{k\lambda^{k-1} y}{k} = \frac{\lambda^k}{k} x + \lambda^{k-1} y$.

Because $|\lambda| = 1$, $\frac{\lambda^k}{k} x$ converges to zero if $k \to \infty$, but $\lambda^{k-1} y$ does <u>not</u> converge to zero if $k \to \infty$, its norm $\|\lambda^{k-1} y\| = \|y\|$ remains constant ($\neq 0$). With an argument that is similar to the proof of theorem (2.3.4-3) in Appendix 2A, it can be shown that in both cases (a) and (b),

$\left\{\frac{CA^{k-1}B}{k}\right\}_{k=1}^{\infty}$ does <u>not</u> converge to zero and

$$\mathrm{tr} \sum_{k=1}^{\infty} \frac{CA^{k-1}BB^*(A^*)^{k-1}C^*}{k^2} = \infty.$$

Page 148, remark (ii): Compare theorem 4.5.2.

Page 155: Add after remark (iii): remark (iv):
Another (and in fact somewhat simpler) representation of (5.2-39)/(5.2 -40) is given in [Hnz 88a,b].

Page 157, section 5.3.2: Δ is the time interval length. This concept corresponds in fact with $\inf \Delta^+$ in definition 2.1.-11.

Page 197 and further: Instead of speaking of a <u>degenerate</u> Riemannian metric, it would probably be better to speak of a positive semi-definite Riemannian metric. If there exists a point Σ and a nonzero tangent vector v, such that at Σ, $g(v,v) = 0$, then one could speak of a <u>singular</u> positive semi-definite Riemannian metric.

Page 201: The last full sentence should read: "One can expect this to hold for all locally continuous canonical forms in which n^2 components of (A,B,C) are fixed to be 0 or 1, and all other components are free to vary in an n(m+m')- dimensional open set, such that its closure contains systems with smaller McMillan degree."

Page 226: Formula (6.1-7) follows from the fact that

$$d\frac{\delta\Gamma}{\delta b} - b\frac{\delta\Gamma}{\delta d} = 0.$$

Page 352: The concept of a limit point at infinity corresponds with that of an ω-limit point of [Hi - S].

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.

59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.

60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict of conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1990.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1990.