Chapter 7

Some Results on Realization Theory

In this chapter, attention will be paid to the question of realizing a state space description for linear systems when only the so-called Markov parameters are given. Not only is this an important theoretical problem, which provides a lot of conceptual insight, but also it is essential in the identification of linear system description from input-output data that are possibly corrupted by noise, which will be the subject of chapter 8. We shall especially be interested in the structural properties of rank deficient block Hankel matrices of Markov parameters and their relation with the singular value decomposition. However, it will turn out that it is not specifically the block Hankel structure which is important, but the so-called *shift structure* of certain subspaces. This will allow to generalize the well known realization theory for block Hankel matrices, to more general matrices, without explicit structure.

This chapter is organised as follows. First, we pay attention to structural properties of rank deficient block Hankel matrices and their relation with singular value decomposition. It is shown how this leads naturally to some well known realization algorithms. Some error bounds are derived for the case of noisy data (section 7.1). In section 7.2. we review some system theoretical aspects of block Hankel matrices, especially concerning their singular value decomposition and the relation to some model reduction strategies. In section 7.3., it is demonstrated how rank deficient Hankel matrices can be parametrized in terms of their singular value decomposition. Use is made of unconventional matrix calculus. Finally, in section 7.4. the derived insights are employed in order to generalize the so-called shift structure of certain subspaces into what we have called structure exploiting factor analysis.

Throughout this chapter, the following notations will be applied for a linear time invariant state space model with m inputs u_k, l outputs y_k and n controllable and observable states x_k:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned}$$

This model will be denoted by the quadruple (A, B, C, D). If B, C and/or D are scalars, vectors, they are replaced by b, c, d. The impulse response matrices of this system are defined

as:

$$H_0 = D \qquad H_k = CA^{k-1}B \;\; k \geq 1$$

For a discrete time system they coincide with the so-called Markov parameters [14]. The transferfunction $H(z)$ of the system equals:

$$H(z) = C(zI - A)^{-1}B + D$$

The extended observability and controllability matrices Γ_k and Δ_k are defined as:

$$\Gamma_k = \begin{pmatrix} C \\ CA \\ \dots \\ CA^{k-1} \end{pmatrix} \qquad \Delta_k = (B \; AB \; \dots \; A^{k-1}B)$$

For $k = n$, these are of course the observability and controllability matrices of the system (A, B, C). Throughout the chapter, it will be assumed that the triplet (A, B, C) is minimal i.e. $rank(\Gamma_n) = rank(\Delta_n) = n$ which is equivalent with the system being both completely observable and controllable.

7.1 Structural properties of rank deficient block Hankel matrices.

In this section, first some properties of general block Hankel matrices are analysed. Next, attention is paid to rank deficient block Hankel matrices and it is shown how the singular value decomposition allows to compute a state space model from the block Hankel matrices of Markov parameters. Furthermore, the system theoretic properties of the singular value decomposition of rank deficient block Hankel matrices are highlighted.

7.1.1 General Block Hankel matrices.

Definition 1 Block Hankel Matrix.
Given a set of $l \times m$ matrices T_k, $k = 1, \dots, K$. The $il \times jm$ block Hankel matrix, with block dimensions i and j and $(i + j - 1) \leq K$, is defined as $H(p, q) = T_{p+q-1}$, where the indices p and q refer to the block indices.

Let us introduce some conventions: If H is a $il \times jm$ block matrix, with $l \times m$ matrices as its blocks, then $\overline{H}$ ($\underline{H}$) is the $(i-1)l \times jm$ matrix constructed from H by omission of the first (last) block row. In the expression $\overline{H}^t$ ($\underline{H}^t$) the first (last) block row is omitted before the transpose is taken. The specific structure of a block Hankel matrix reveals itself in any possible factorization of the matrix:

Theorem 1 Factorization structure of Hankel matrices of finite dimensions
Let H be a $il \times jm$ matrix that can be factored as

$$H = XY^t$$

where X and Y are arbitrary matrices. Then H is a block Hankel matrix if and only if

$$\overline{X}\underline{Y}^t = \underline{X}\overline{Y}^t$$

Proof: Follows immediately from the block Hankel structure. □

7.1.2 Block Hankel matrices with Markov parameters.

Rank deficient block Hankel matrices play an important role in the analysis and realization of linear systems. Although a lot of their algebraic properties have been known for quite a long time [12] it is only since the introduction of the now well known realization scheme of Ho - Kalman [11] that their system theoretic importance has been established. An important step was the introduction in the realization context of the singular value decomposition of the Hankel matrix of Markov parameters [25]. Moreover there exists a close connection between the singular value decomposition and the concept of balanced realization, introduced in [18] where it was demonstrated how the singular values and vectors of Hankel matrices of finite and infinite dimensions can be used as quantitative measures of controllability and observability. This connection is exploited in [9] to solve the celebrated model reduction problem with Hankel norm using state space methods only.

A fundamental tool in the analysis of linear systems is the block Hankel matrix, constructed from the Markov parameters of the linear system :

$$H(p,q) = H_{p+q-1} = CA^{p+q-2}B \quad p = 1,\ldots,i \quad q = 1,\ldots,j$$

The following properties of this block Hankel matrix can be considered as classical in realization theory of linear systems :

Lemma 1 Factorization property of Hankel matrices from Markov parameters
Let $H_k = CA^{k-1}B$ be the Markov parameters of a linear system and $H(p,q) = H_{p+q-1}$ be a $il \times jm$ block Hankel matrix . Then

$$H = \Gamma_i \Delta_j$$

Proof: Trivial. □

Lemma 2 Rank property.
Let $H_k = CA^{k-1}B$ be the Markov parameters of a linear system and $H(p,q) = H_{p+q-1}$ be a $il \times jm$ block Hankel matrix with $i \geq n$ and $j \geq n$. Then $rank(H) = n$.

Proof: Follows from lemma 1 and the minimality of (A,B,C). □

The second lemma indicates that the rank of a block Hankel matrix of Markov parameters is always equal to the dimension of the observable and controllable part of the state space, if the block dimensions i and j are chosen large enough ($i, j \geq n$). A unique correspondance between a linear system defined by a transfer function $H(z)$ and a block Hankel matrix only exists for block Hankel matrices with infinite dimension. Indeed, the transfer function $H(z)$ is uniquely represented by the corresponding infinite sequence of Markov parameters. When dealing with the correspondance of a finite dimensional block Hankel matrix of a linear system $H(z)$, uniqueness of the realization can only be achieved by posing extra conditions on the extension of the sequence of Markov parameters to infinity, e.g. the extra condition of minimality of the partial realization of the finite sequence. As a result, the problem of minimal partial realization is involved [4].

Let H be a $il \times jm$ block Hankel matrix of rank n where $n < i$ and $n < j$. Define H_1 to be the $(i-1)l \times jm$ submatrix of H consisting of its first $i-1$ block rows and H_2 the $il \times (j-1)m$ submatrix of H consisting of its first $j-1$ block columns.

Lemma 3 *If $rank(H) = rank(H_1) = rank(H_2) = n$, then the blocks of H are Markov parameters of a linear system of minimal state space dimension n.*

Proof: see e.g. [4]. □

Note that the condition of rank deficiency of a finite block Hankel matrix is not enough to guarantee a unique solution to the corresponding minimal realization problem. The necessary and sufficient condition hereto is the rank condition on the submatrices H_1 and H_2. This criterion is well known in the analysis of the minimal partial realization property as the *partial realizability criterion*:

$$rank(H) = rank(H_1) = rank(H_2) = n \;\; with \;\; i, j > n$$

It will be assumed throughout that this partial realizability criterion is satisfied.
The matrix $(\overline{H} \;\; \underline{H})$ is a $(i-1)l \times (2jm)$ matrix constructed by concatenation of $\overline{H}$ and $\underline{H}$. The singular values of a matrix are the most reliable measures to estimate the rank of a matrix. Moreover, the singular value decomposition provides a factorization of the block Hankel matrix so that the following property is easy to prove.

Corollary 1 *Let H be a $il \times jm$ matrix. Let the SVD of H be $H = USV^t$. Then H is a block Hankel matrix if and only if $\overline{U}S\underline{V}^t = \underline{U}S\overline{V}^t$*

Proof: Choose for X and Y in theorem 2: $X = U$ and $Y = VS$. □

Hence the structure of the block Hankel matrix reveals itself in the SVD. The specific structure of rank deficient block Hankel matrices will be translated into a condition on the rank of some concatenated matrices that are constructed from the singular vectors.

Theorem 2 *Let H be a $il \times jm$ block Hankel matrix of rank n with $i > n$ and $j > n$ and SVD $H = USV^t$ where S is square $n \times n$. Assume that the submatrices H_1 (first $i-1$ block rows) and H_2 (first $j-1$ block columns) satisfy the partial realization criterion:*

$$rank(H) = rank(H_1) = rank(H_2) = n$$

Then:

$$rank(\overline{U} \;\; \underline{U}) = rank(\overline{V} \;\; \underline{V}) = n$$

Proof: From corollary 1:

$$(\underline{U} \;\; \overline{U}) \begin{pmatrix} S & 0 \\ 0 & -S \end{pmatrix} \begin{pmatrix} \overline{V}^t \\ \underline{V}^t \end{pmatrix} = 0$$

Now define $s = rank(\underline{U} \;\; \overline{U})$. From the SVD of H it follows that $H_1 = \underline{U}SV^t$ and from the rank condition $rank(H_1) = n$, it follows that $rank(\underline{U}) = n$. Hence, $s \geq n$. Then:

$$dim \;\; ker \;\; \left((\underline{U} \;\; \overline{U}) \begin{pmatrix} S & 0 \\ 0 & -S \end{pmatrix} \right) = 2n - s$$

Now define $t = rank(\overline{V}\ \ \underline{V})$. From the SVD of H it follows that $H_2 = US\underline{V}^t$ and from the condition $rank(H_2) = n$, it follows that $rank(\underline{V}) = n$. Hence , $t \geq n$. Since the columnspace of the matrix $(\overline{V}\ \ \underline{V})^t$ belongs to the kernel of the matrix

$$\left((\underline{U}\ \ \overline{U})\begin{pmatrix} S & 0 \\ 0 & -S \end{pmatrix}\right)$$

we have that $(2n-s) \geq t \geq n$ so that both $(2n-s) \geq n$ (or $s \leq n$) and $s \geq n$ have to be satisfied. Hence $s = n$ or $rank(\underline{U}\ \ \overline{U}) = n$. The proof for $rank(\underline{V}\ \ \overline{V}) = n$ is of course similar. □

An immediate consequence of theorem 2 is the special shift structure of the left and right singular vectors of the Hankelmatrices, which is well known and explored in realization theory to obtain a state space model from the Markov parameters [11], [16], [25]:

Corollary 2 *Let H be a $il \times jm$ block Hankel matrix of rank n with $i > n$ and $j > n$ and SVD $H = USV^t$ where S is an $n \times n$ diagonal matrix with the singular values. Then U and V have the following structure (which will be called shift structure):*

$$U = \begin{pmatrix} u \\ uT_u \\ uT_u^2 \\ \cdots \\ uT_u^{i-1} \end{pmatrix} \qquad V = \begin{pmatrix} v \\ vT_v \\ vT_v^2 \\ \cdots \\ vT_v^{j-1} \end{pmatrix}$$

where u is a $i \times n$ matrix, v is a $j \times n$ matrix and T_u and T_v are square $n \times n$ matrices which are the unique solution of $\underline{U}T_u = \overline{U}$, $\underline{V}T_v = \overline{V}$. Moreover $ST_v = T_uS$.

The matrices T_u and T_v will be called shift matrices.

Proof: From theorem 2 and the partial realization condition, it follows that there must exist unique square $n \times n$ matrices T_u and T_v such that $\underline{U}T_u = \overline{U}$ and $\underline{V}T_v = \overline{V}$. T_u and T_v are given by:

$$T_u = (\underline{U}^t\underline{U})^{-1}\underline{U}^t\overline{U} \qquad T_v = (\underline{V}^t\underline{V})^{-1}\underline{V}^t\overline{V}$$

If u and v are the first block row of U and V, the specific shift structure follows from $\underline{U}T_u = \overline{U}$ and $\underline{V}T_v = \overline{V}$. From corollary 1: $\underline{U}S\overline{V}^t = \overline{U}S\underline{V}^t$ hence:

$$S\overline{V}^t\underline{V}(\underline{V}^t\underline{V})^{-1} = (\underline{U}^t\underline{U})^{-1}\underline{U}^t\overline{U}S$$

so that $ST_v = T_uS$. □

Corollary 2 specifies completely how to obtain a state space realization of a set of Markov parameters of a discrete system if the partial realization condition is satisfied. The eigenvalues of the shift matrices T_u and T_v are of course the controllable and observable poles of the system. An important remark is the fact that the shift structure is a property of the columnspace of the matrices U and V and not of the specific choice of basis in that space:

Lemma 4 *Let U be a $il \times n$ matrix $(i > n)$ of rank n with shift structure:*

$$U = \begin{pmatrix} u \\ uT_u \\ uT_u^2 \\ \cdots \\ uT_u^{i-1} \end{pmatrix}$$

satisfying rank$(\underline{U}) = n$ and let P be any $n \times n$ nonsingular matrix. Then the matrix (UP) is also a matrix with shift structure with a shift matrix that is similar with T_u.

Proof: Since $\underline{U}T_u = \overline{U}$, it follows that $\underline{U}T_uP = \overline{U}P$ and because P is nonsingular we have that $(\underline{U}P)(P^{-1}T_uP) = (\overline{U}P)$. Hence the matrix (UP) has shift structure with shift matrix $P^{-1}T_uP$ which is similar to T_u and hence has the same Jordan structure. □

The interpretation of lemma 4 is straightforward: With the columnspace of a matrix U with shift structure, one can associate a set of eigenvalues that only depend on that vectorspace and not on the choice of basis in that space.

7.1.3 Realization Algorithms.

Realization theory is concerned with the determination of a state space model, given the Markov parameters of the system. We shall briefly review two SVD based algorithms, and then present a deductive error analysis for both of them. The first one is the Zeiger-McEwen algorithm [25] while the second one is the approach proposed by Kung [16]. While historically, the Zeiger-McEwen approach was the first, the Kung algorithm is more interesting from the point of view of our identification approach presented in chapter 8. The Kung approach explicitly exploits the close connection between a subspace, its shift structure and the minimal set of eigenvalues.

The algorithm of Zeiger - McEwen

Given at least $i + j$ Markov parameters $H_1, \ldots, H_{i+j}$ of a linear system with minimal state space dimension n where $i > n$ and $j > n$.

1. Construct two block Hankel matrices H^1 and H^2 from the Markov parameters with block dimensions i and j. H^1 contains the Markov parameters $H_1, \ldots, H_{i+j-1}$ while H^2 contains $H_2, \ldots, H_{i+j}$.

2. Consider the singular value decomposition of H^1:

   $$H^1 = USV^t$$

   Here U and V are $li \times n$ and $mj \times n$ orthonormal matrices and S is a square diagonal $n \times n$ matrix where n is the minimal dimension of the system.

3. If the matrix S is rewritten as $S = S^{1/2}S^{1/2}$, a realization for the matrices B and C is easily obtained as:

   $$C = U(1:l, 1:n)S^{1/2} \qquad B = S^{1/2}(V(1:m, 1:n))^t$$

   where the notation $(1:l)$ denotes the index range for rows and columns.

4. The matrix A is obtained as:

$$A = S^{-1/2} U^t H^2 V S^{-1/2}$$

The proof of the validity of this algorithm for exact data is straightforward using the previously derived observations on the block Hankel structure.

The algorithm of Kung

While the Zeiger-McEwen approach requires the construction of 2 block Hankel matrices, the algorithm proposed in [16] only requires one:

1. Construct the block Hankel matrix H with Markov parameters and block dimensions i and j.

2. From its singular value decomposition:

$$H = USV^t$$

obtain the matrices B and C as:

$$C = U(1:l, 1:n)S^{1/2} \quad B = S^{1/2}(V(1:m, 1:n))^t$$

3. The matrix A follows from:

$$\underline{U} S^{1/2} A = \overline{U} S^{1/2}$$

or:

$$A = S^{-1/2}(\underline{U}^t \underline{U})^{-1} \underline{U}^t \overline{U} S^{1/2}$$

Note that the partial realization criterion ensures that the matrix $(\underline{U}^t \underline{U})$ is invertible. It is interesting to note that sometimes in the literature, the algorithm of Kung is misunderstood: additional zeros are introduced in order to compute the matrix A from the shift structure of the block Hankel matrix. This is of course completely unnecessary and leads to a very bad implementation of Kung's approach, especially for signals with almost no damping.
Of course, in the case of noisy data, the computation of the matrix A corresponds to a least squares estimate and could be replaced by a total linear least squares estimate instead, since both the left and right hand side are obtained from the same matrix and hence are corrupted by comparable noise levels. However, via a matrix inversion lemma, it is possible to reduce the computational complexity of the inversion in case of least squares. Assume that u is the last block row of U, consisting of l rows. Then from the orthonormality of U it follows that:

$$\underline{U}^t \underline{U} = I_n - u^t u$$

Hence:

$$(\underline{U}^t \underline{U})^{-1} = I_n + u^t (I_l - uu^t)^{-1} u$$

Hence, it is only needed to invert a $l \times l$ matrix. For instance, if $l = 1$, only the inverse of a scalar is to be computed. However, observe that the conditioning of the inverse depends upon the singular values of u. When some of them are close to one, the inversion will be ill-conditioned. It can be proved however, that the partial realization criterion guarantees that the maximal singular value of u is smaller than 1 [5].

Other algorithms

Variations on the above algorithms can be found in [3]. They are based upon the observation that the block Hankel structure is not a *conditio sine qua non* in order to take profit from the shift structure. As a matter of fact, also the so-called page matrix satisfies a shift structure:

$$\begin{pmatrix} H_1 & H_2 & \dots & H_j \\ H_{i+1} & H_{i+2} & \dots & H_{2j} \\ \dots & \dots & \dots & \dots \\ H_{(i-1)j+1} & H_{(i-1)j+2} & \dots & H_{ij} \end{pmatrix} = \begin{pmatrix} C \\ CA^j \\ \dots \\ CA^{(i-1)j} \end{pmatrix} (B \;\; AB \;\; \dots \;\; A^{j-1}B)$$

A careful analysis is needed in order to investigate under which circumstances observability is preserved for the model. However, observe that the shift structure will be present in the singular value decomposition in the case of exact data. Comparisons between the block Hankel matrix approach and the page matrix approach can be found in [3].

Sensitivity analysis for noisy data.

When the data are exact, all realization algorithms deliver the same result. If however the data are noisy, a deductive analysis is necessary in order to provide necessary error bounds on the obtained results.

An interesting error bound has been obtained in [5]. It provides a better estimate than the one derived by Kung in [16].
Denote the 'exact', pure noise and noisy block Hankel matrix by $\hat{H}$, $\tilde{H}$ and H, the additivity of the noise model allows to state that:

$$H = \hat{H} + \tilde{H}$$

Generically, the rank of H will be larger than the rank of $\hat{H}$. Both the Zeiger-McEwen and the Kung approach replace H by an approximation, based on its singular value decomposition:

$$H = USV^t = (U_1 \;\; U_2) \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_1^t \\ V_2^t \end{pmatrix}$$

where S_1 contains the n largest singular values. The matrix H is then approximated by its least squares approximation $U_1 S_1 V_1^t$, which has however no block Hankel structure. The following theorem assesses the relative accuracy of the estimate A of the exact system matrix $\hat{A}$ for both algorithms, if, despite the absence of any shift structure, the algorithms are applied formally:

Theorem 3 *For both the Zeiger-McEwen and the Kung algorithm, there exists a similarity transformation T and a constant γ such that:*

$$\|A - T\hat{A}T^{-1}\| < \gamma \frac{(\|\tilde{H}\| + \sigma_{n+1})}{(\sigma_n - \sigma_{n+1})}$$

where the σ_i are the singular values of H, provided that $(\|\tilde{H}\| + \sigma_{n+1}) < 2(\sigma_n - \sigma_{n+1})$.

Proof: see [5]. □

7.2 System theoretic aspects of block Hankel matrices.

In this section, we shall pay attention to certain system theoretic aspects of block Hankel matrices of Markov parameters of multivariable systems. We shall limit ourselves to the results that are well describable in terms of the singular value decomposition, hence emphasizing both the conceptual as the computational advantages of this matrix decomposition technique.

7.2.1 The block Hankel matrix as interface between past and future.

The block Hankel matrix can be considered as the matrix of the linear transformation between past inputs and future outputs of a system in the following sense. Consider an input sequence in the past:

$$u^- = \begin{pmatrix} u_{-1} \\ u_{-2} \\ \cdots \\ u_{-j} \end{pmatrix}$$

Then, obviously, the present state equals $x_0 = \Delta_j u^-$. If from time 0 on, the input sequence is zero, the observed output will be given by:

$$y^+ = \begin{pmatrix} y_0 \\ y_1 \\ \cdots \\ y_{i-1} \end{pmatrix} = \Gamma_i x_0 = \Gamma_i \Delta_j u^- = H u^-$$

This input-output result implies that the singular values of the block Hankel matrix characterize in a quantitative way the input-output properties of the system. As a matter of fact, it will be shown furtheron how the singular values of the block Hankel matrix are important with respect to controllability, observability and model reduction. Observe that the state is the interface between past and future. As a matter of fact, this qualification can be restated into a rigorous definition of the notion of a state, even for non-linear and time-varying systems [14]. It will be the basis of some identification strategies in chapter 8.

7.2.2 Oriented energy and controllability and observability.

In this section, we shall investigate quantitative measures of the degree of controllability and observability of a linear system. The singular value decomposition turns out to be a very powerful tool. Controllability and observability are qualitative characterizations of systems, that were first described in the context of state computation (Kalman filter) and optimal control. However, only recently, one started to be interested in more quantitative characterizations in order to be able to answer questions like: *How* controllable (observable) is a system? Since both controllability and observability can be investigated by verifying the rank of certain matrices, it must come as no surprise that once again, a more quantitative description is possible in terms of the singular values of the same matrices. The results will allow to get insight into certain design questions of linear systems and into the rationale behind model reduction. Moreover, the interpretation turns out to be important with respect to control strategies and the identification algorithms to be derived in chapter 8.

Controllability

Somewhat loosely speaking, controllability is that property of a (linear) system, which states that it is possible to reach each specified state in the ambient state space. Assuming that the initial state is x_0, it is straightforward to show that at time k the state is given by:

$$\begin{aligned} x_k &= A^k x_0 + A^{k-1}Bu_0 + A^{k-2}Bu_1 + \ldots + Bu_k \\ &= A^k x_0 + [\, B\ AB\ \ldots\ A^{k-1}B\,] \begin{pmatrix} u_k \\ u_{k-1} \\ \ldots \\ u_1 \\ u_0 \end{pmatrix} \end{aligned}$$

or in shorthand notation:

$$x_k - A^k x_0 = \Delta_k U$$

Observe that for $k = n$, Δ_k is the so called controllability matrix of the system. If this matrix is of full rank n, the system is qualified as controllable. Let's analyse the preceding equations in a more geometrically inspired manner. Observe that the vector $x_k - A^k x_0$ is obtained from the columns of Δ_k via linear combinations of those columns, with weights provided by the inputs. Hence, only those states can be reached for which:

$$(x_k - A^k x_0) \in span_{col}(\Delta_k)$$

Now let's have a look at the following problem:

Find the input of minimum energy that brings the system from initial state x_0 to end state x_k at time k.

Obviously, this problem has only a solution if the above condition is satisfied. It is assumed that this is the case (otherwise one should consider least squares approximations but this will not be considered here.) The solution follows immediately from a well known minimum norm property of the singular value decomposition in the solution of underdetermined sets of linear equations. Hereto consider the singular value decomposition of Δ_k:

$$\Delta_k = U_{\Delta_k} \Sigma_{\Delta_k} (V_{\Delta_k})^t$$

where U_{Δ_k} is a $n \times r_k$, Σ_{Δ_k} is a $r_k \times r_k$, and V_{Δ_k} is an $mk \times r_k$ matrix, r_k being the rank of Δ_k. Then the optimizing input sequence is given by:

$$U_{opt} = V_{\Delta_k} (\Sigma_{\Delta_k})^{-1} (U_{\Delta_k})^t (x_k - A^k x_0)$$

Recall the condition that $(x_k - A^k x_0) \in span_{col}(\Delta_k)$. Especially for $k < n$ this condition may be restrictive. It is well known that the rank of Δ_k is non-decreasing as a function of k going to n. Those values of k for which the rank increases, are the so called *controllability indices* of the system and as can be now easily seen, they play a fundamental role in minimum time and minimum energy control strategies. However, for $k \geq n$, the rank of Δ_k equals that of Δ_n. This follows directly from the Cayley-Hamilton theorem.

The mimimum input 'energy' that is needed to control the state from x_0 to x_k, can now be computed. Hereto, call $x_{0k} = x_k - A^k x_0$.

$$\begin{aligned} J_{min} &= \sum_{i=0}^{k} u_i^t u_i \\ &= x_{0k}^t U_{\Delta_k} \Sigma_{\Delta_k}^{-1} (V_{\Delta_k})^t V_{\Delta_k} (\Sigma_{\Delta_k})^{-1} (U_{\Delta_k})^t x_{0k} \\ &= (x_{0k}^t U_{\Delta_k}) \begin{pmatrix} 1/\sigma_1^2 & & & \\ & 1/\sigma_2^2 & & \\ & & \cdots & \\ & & & 1/\sigma_n^2 \end{pmatrix} (U_{\Delta_k})^t x_{0k}) \\ &= x_{0k}^t P_k^{-1} x_{0k} \end{aligned}$$

where P_k is the so called controllability Grammian $P_k = \Delta_k \Delta_k^t$. Obviously, the oriented energy distribution of the vector sequence of columns of Δ_k is the key tool to obtain quantitative measures of the degree of controllability. If for instance x_{0k} equals the first left singular vector, then the minimum energy needed to control x_0 to x_k, equals $1/\sigma_1^2$. If however x_{0k} equals the smallest singular vector, the energy needed for control equals $1/\sigma_n^2$. Hence, states vectors that are lying in the direction of the stronger singular vectors are relatively easy to control compared to states that are lying in small singular subspaces. These require a lot of energy to reach them. In the limit case, one or more of the singular values of the controllability matrix are zero. This corresponds to directions in the state space that are not controllable. Observe that as a global measure of controlllability, one could take the condition number of the controllability matrix. If this condition number is large, then there exist states that are difficult to control with respect to others. If however the condition number is close to its minimum value one, then all states are controllable with the same required minimum energy.

As a final remark, observe that the presented derivation is not reference frame independent. Indeed, the controllability matrix and hence its singular value decomposition, explicitly depend upon the choice of basis in the state space. Once such a basis has been fixed, the previous discussion applies, *with respect to that specific basis.* Observe that the qualification of controllability (i.e. controllable or not) is independent of the basis, but that its quantification (how controllable) is determined by the specific basis choice. However, we shall show furtheron that certain bases are to be preferred, namely the so called *balanced coordinates.* As an extreme example of how a non-singular $n \times n$ transformation T might affect the oriented energy distribution that characterizes the degree of controllability, simply choose $T = (\Sigma_{\Delta_k})^{-1}(U_{\Delta_k})^t$ which modifies the controllability matrix Δ_k into $T\Delta_k = (V_{\Delta_k})^t$. It is not difficult to show now, that with respect to this coordinate system, the oriented energy distribution is isotropic. Hence, all states are equally well controllable.

Observability

Since the qualification of observability is the system theoretic *dual* of controllability, it comes as no surprise that the quantitative characterization of observability will follow the same lines of thought as that for controllability. To be more specific, consider the autonomous linear

system:

$$\begin{aligned} x_{k+1} &= Ax_k \\ y_k &= Cx_k \end{aligned}$$

and assume that this systems starts from an initial state x_0. It is straightforward to show that:

$$y_k = CA^{k-1}x_0$$

and that:

$$\begin{pmatrix} y_0 \\ y_1 \\ \cdots \\ y_{k-1} \end{pmatrix} = \begin{pmatrix} C \\ CA \\ \cdots \\ CA^{k-1} \end{pmatrix} x_0$$

This becomes in short hand notation:

$$y = \Gamma_k x_0$$

Very much the same way as controllability indices are defined, the same can be done for observability indices. They are important in the derivation and analysis of filtering problems, where one aims at computing optimally and this as soon as possible, the state. It is also easily seen that, if the initial state belongs to the row space of the observability matrix, it can be computed as the solution of a set of linear equations. If however it does not completely belong to this row space, the component of it, orthogonal to the row space of Γ_k, will be unobservable. Hence, the part of the ambient n-dimensional state space that is observable, is precisely $span_{row}(\Gamma_n)$. From now on, it is assumed that:

$$x_0 \in span_{row}(\Gamma_k)$$

Let's now invertigate the energy contribution of an initial state x_0 in the output sequence. As a measure of the output energy, it is quite natural to take the following sum:

$$\begin{aligned} J_y &= \sum_{i=0}^{k} y_i^t y_i \\ &= \sum_{i=0}^{k} x_0^t ((A^i)^t C^t C A^i) x_0 \\ &= x_0^t Q_k x_0 \end{aligned}$$

Here, Q_k is the so called observability Grammian: $Q_k = (\Gamma_k)^t \Gamma_k$. If the singular value decomposition of Γ_k is given by:

$$\Gamma_k = U_{\Gamma_k} \Sigma_{\Gamma_k} (V_{\Gamma_k})^t$$

then the output energy can be written as:

$$J_y = (x_0^t V_{\Gamma_k})(\Sigma_{\Gamma_k})^t (\Sigma_{\Gamma_k})((V_{\Gamma_k})^t x_0)$$

Obviously, this expression can again be interpreted in terms of oriented energy. If the initial state has an important component along the smallest singular direction, it will represent a

small energy contribution in the output. If however, the initial state is lying along the largest singular vector, then it will contribute a lot in the output energy.

Finally, observe that the same remarks apply as for the controllability analysis: the condition number of the observability matrix Γ_k is a good measure for the overall observability but once again, the quantification represented by the singular values, heavily depends upon the specific coordinate system of the state space.

7.2.3 Choosing a state space basis

Both in the discussion of the relation between oriented energy and the properties of controllability and observability, it was observed that the singular values depend heavily on the specific choice of the coordinate system in the state space. Of course, this represents a serious drawback for this kind of analysis since it is a most desirable feature of a scientific theory that it is coordinate frame free, i.e. its properties must be independent upon the specific choice of references.

It will now be argumented how, in some sense, the singular value decomposition of the block Hankel matrix containing the Markov parameters $H_i = CA^{i-1}B$, provides a specific realization of the system, the so called *balanced realization.* Since its discovery, this specific realization has been proved to be superior with respect to others, and this because of its numerical reliability and insensitivity to finite precision arithmetic ([19], [23], ...). In addition, it has recently be proved, that the (by now classical) Adamjan-Arov-Krein approach for model reduction can be derived completely in a state space formalism, if the balanced realization of the system is used. The model reduction criterion that is optimized is the minimization of the maximal singular value of the difference between the original block Hankel matrix and the one of the reduced model [9].

The singular value decomposition of the block Hankel matrix with the Markov parameters also provides a rationale for model reduction in more energy oriented norms and measures, that will prove to be extremely efficient in the identification schemes to be derived in chapter 8.

In this work, we shall use a slight generalization of the balanced realization concept as originally derived in [18]. In the original formulation, the block Hankel matrices that are considered are double infinite (infinite number of rows and columns), hence restricting the results necessarily to *stable* linear systems. There is however no reason to limit ourselves to stable systems, as long as the block Hankel matrices are restricted to be finite (and the dynamic range of the numerical precision tolerates the used dimensions). The definition of the balancedness of a system representation will now be done directly in terms of the singular value decomposition of the block Hankel matrix, though one could follow other paths of derivation [18].

Previously, it was proved that any factorization of the block Hankel matrix with the Markov parameters, provides a valid controllability and observability matrix. Since the singular value decomposition provides a very priviliged decomposition, it may be expected that the corresponding observability and controllability matrices are special too. This is indeed the case as will now be shown.

Theorem 4 *Let H be a block Hankel matrix with i block columns and j block columns, containing the $l \times m$ Markov parameters of a system with minimal (= controllable and observable) dimension n. Assume that $i > n$ and $j > n$. Let the singular value decomposition of H be given by:*

$$H = U_H S_H V_H^t$$

where U_H is the $il \times n$ matrix with the left singular vectors, Σ is the $n \times n$ matrix with the singular values and V is a $jm \times n$ matrix with the right singular vectors. Then $U_H S_H^{\alpha}$ is an observability matrix and $S_H^{1-\alpha} V_H^t$ is a controllability matrix.

Proof: Trivial. See also [6] □

Despite the somewhat trivial contents of the theorem, it provides a considerably powerful coordinate system in the state space. The corresponding expressions for the system matrices, constitute the balanced realization.

Definition 2 Balancedness of a realization
A realization is called balanced if the controllability and observability matrices have the forms of a singular value decomposition as in theorem 4, with the specific choice for $\alpha = 1/2$.

Observe that, as long as i and j remain finite, the definition also applies for unstable systems. We like to emphasize this fact because in a lot of (mainly control) applications, the system under consideration is unstable. Originally, balancedness was defined for stable systems only, in terms of the singular value decomposition (principal component analysis) of a double infinite Hankel matrix [18]. Our definition allows the treatment of unstable system via a restriction to finite dimensions. A property of the balancedness of a realization is that the corresponding controllability and the observability Grammians are diagonal and equal to each other, being the diagonal matrix with the squared singular values of the Hankel matrix of Markov parameters (of course, this qualification can also be taken as a definition of balancedness). If $P_k = \Delta_k \Delta_k^t$ and $Q_k = \Gamma_k^t \Gamma_k$, then it is an easy exercise to demonstrate that they satisfy the following Lyapunov functions:

$$\begin{aligned} AP_kA^t - P_k &= A^{k+1}BB^t(A^{k+1})^t - BB^t \\ A^tQ_kA - Q_k &= (A^{k+1})^tC^tCA^{k+1} - C^tC \end{aligned}$$

If there is no pair of eigenvalues λ_i, λ_j of A such that $\lambda_i \lambda_j = 1$ (with possibly $i = j$), then these solutions are unique. When the system is stable, taking the limit for $k \to \infty$ allows to obtain the following Lyapunov equations for the infinite Grammians:

$$\begin{aligned} AP_\infty A^t - P_\infty &= -BB^t \\ A^tQ_\infty A - Q_\infty &= -C^tC \end{aligned}$$

A method to obtain a balanced realization from the solution of the set of Lyapunov equations from an arbitrary realization of the system can be found in [2].
The specific coordinate system provided by the singular value decomposition of the block Hankel matrix with Markov parameters, thanks its qualification of *balancedness* to two reasons:

- In this specific coordinate system, the quantitative analysis of both controllability and observability as discussed in section 7.2.2., can be done *in terms of the same set of singular values*, namely those of the block Hankel matrix. This can be easily verified by inspection of the specific forms of Γ_k and Δ_k. As a matter of fact, while the singular values of Γ_k and Δ_k are determined by the specific state space basis, the singular values of the product $\Gamma_k\Delta_k$ are *input-output invariants*, with possibly interpretations in the frequency domain as well [9].

- A second reason is the fact that the singular values of this specific coordinate system, provide an excellent rationale for model reduction, based upon energy considerations. One can perform the input-output analysis via the block Hankel matrix, hereby paying equally well attention to the controllability and observability aspects of the system. This is expressed by the choice of $\alpha = 1/2$. Other possibilities which might be of interest in certain applications are $\alpha = 1$ (input normal) and $\alpha = 0$ (output normal)[24].

- Observe that all these results, are essentially *open loop* characterizations. When the system operates in closed loop, a singular value analysis and the corresponding balanced model reduction procedure, can deliver misleading results. The closed loop analysis is much more complicated and exploits the invariance of the eigenvalues of the products of the solutions to the optimal filtering and control Ricati equations, as was developed in the interesting article of Jonckheere and Silverman [13].

The following derivation provides (an at least heuristic) rationale for the use of the balanced coordinates in model reduction applications. Without loss of generality, it will be assumed that the block Hankel matrix with Markov parameters has the same number of block columns as block rows and that the observability and controllability matrix are computed from its SVD:

$$\begin{aligned} H &= \begin{pmatrix} C \\ CA \\ \cdots \\ CA^{k-1} \end{pmatrix} (B\ AB\ \ldots\ A^{k-1}B) \\ &= \Gamma_k\Delta_k \\ &= (US^{1/2})(S^{1/2}V^t) \end{aligned}$$

Hence, the state space representation is balanced. It is assumed that the matrix S is square diagonal $n \times n$.

Model reduction and controllability.

Assume that the singular values are partitioned into strong and weak ones:

$$S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$$

where S_1 contains the r largest and S_2 the $n - r$ smallest singular values. Starting from an initial state $x_0 = 0$, the state at time instant x_k is given by:

$$x_k = \Gamma_k \begin{pmatrix} u_{k-1} \\ u_{k-2} \\ \cdots \\ u_0 \end{pmatrix} = S^{1/2}V^t u$$

When the state vector and the right singular matrix are partioned according to the partitioning of the singular values, one finds that:

$$x_k = \begin{pmatrix} x_k^1 \\ x_k^2 \end{pmatrix} = \begin{pmatrix} S_1^{1/2} V_1^t \\ S_2^{1/2} V_2^t \end{pmatrix} u$$

Now assume that $u^1(.)$ is the input sequence that brings the system from $x_0 = 0$ to the state:

$$\begin{pmatrix} x_k^1 \\ 0 \end{pmatrix}$$

and that $u^2(.)$ is that input sequence that brings the system from $x_0 = 0$ to the state:

$$\begin{pmatrix} 0 \\ x_k^2 \end{pmatrix}$$

Then, it is easy to see that:

$$max_{over\ all\ u^1,u^2} \frac{\|x_k^2\|^2}{\|x_k^1\|^2} = \frac{max_{u^2}\ \|S_2^{1/2} V_2^t u^2\|^2}{min_{u^1}\ \|S_1^{1/2} V_1^t u^1\|^2} = \frac{\sigma_{r+1}}{\sigma_r}$$

Hence it holds that:

$$\|x_k\|^2 = \|x_k^1\|^2 + \|x_k^2\|^2 = \|x_k^1\|(1 + \frac{\|x_k^2\|^2}{\|x_k^1\|^2}) \leq \|x_k^1\|^2(1 + \frac{\sigma_{r+1}}{\sigma_r})$$

Hence, when expressed in balanced coordinates, most of the energy of the state vector is caused by the part associated to the largest singular values of the block Hankel matrix of Markov parameters. This is the rationale behind open loop model reduction via singular value decomposition, from the point of view of control. If the gap between the singular values σ_r and σ_{r+1} is large, there is very little information lost about the state.

Model reduction and observability.

Assume that the initial state of the system is x_0 and that the system operates autonomously, i.e. $u(.) = 0$. Then the output can be obtained from:

$$y = \begin{pmatrix} y_0 \\ y_1 \\ \cdots \\ y_k \end{pmatrix} = \Gamma_k x_0 = U S^{1/2} x_0$$

Applying the same partitioning to the vector x_0 and the left singular matrix U, one finds:

$$y = y_1 + y_2 = U_1 S_1^{1/2} x_0^1 + U_2 S_2^{1/2} x_0^2$$

It is straightforward to verify that:

$$max_{over\ all\ x_0^1,x_0^2} \frac{\|y_2\|^2}{\|y_1\|^2} = \frac{\sigma_{r+1}}{\sigma_r}$$

Hence:

$$\|y\|^2 = \|y_1\|^2 + \|y_2\|^2 = \|y_1\|^2(1 + \frac{\|y_2\|^2}{\|y_1\|^2}) \leq \|y_1\|^2(1 + \frac{\sigma_{r+1}}{\sigma_r})$$

Hence, whenever $\sigma_r \gg \sigma_{r+1}$, that part of the state that corresponds to the largest singular values, when the state is expressed in balanced coordinates, causes most of the energy in the output.

The combination of these controllability and observability properties of the balanced realization, provides a rationale for model reduction of systems, when one is interested in an approximation of the input-output behavior of the system. Results concerning the stability of the reduced system if the original one is stable, can be found in [21]. Recently, it was demonstrated that the singular values *alone* do not provide sufficient information for model reduction. One needs additional characteristics which are called the balanced gains [15]. Other interesting applications of the balanced realization include the optimization and analysis of round off properties of digital filters [19][20], the solution in terms of balanced state space models of the model reduction in Hankel norm [9] and high resolution spectral analysis methods [5] [6] [17].

7.3 Shift Structure and SVD

7.3.1 The SVD of rank deficient block Hankel matrices.

In this section, some new numerical-algebraic results will be established that describe the link between the block Hankel structure and the singular value decomposition. The main result is a parametrization of all rank deficient Hankel matrices of finite dimensions that have a prescribed set of minimal system poles. Use is made of the matrix calculus of Kronecker and Khatri-Rao products (appendix B). The motivation for this research arises from the important relation between the singular values of the Hankel matrix and the degree of controllability and observability of a system described in the preceding section. If for a system of minimal order n, the smallest singular value is small with respect to the largest one, then there exist states that are difficult to control and to observe in terms of 'energy'. Hence, in the design of linear systems, it might be interesting to optimize the block Hankel matrix, which is the product of controllability and observability matrices of the system, such that it is as orthonormal as possible. In this way, the energy to control and observe states may be minimized while the desired minimal system poles are specified (e.g. pole placement).

The key observation that will be exploited is lemma 4 which states that the minimal system poles depend only on the vectorspace spanned by the left and right singular vectors of the rank deficient block Hankel matrix of Markov parameters. The parametrization will be obtained by considering all possible bases in those vectorspaces and yet still respecting the block Hankel structure. Hence we shall obtain some remarkable relations between the block Hankel structure and the singular value decomposition.

We shall intensively use the properties of the Kronecker and the Khatri-Rao product, for which the interested reader is referred to appendix B. The Kronecker product of two matrices is represented by the symbol $\otimes$ while the Khatri-Rao product is symbolized by $\odot$. $vec(A)$ stacks the columns of a matrix A in a long columnvector while $vecd(A)$ does the same, but only

with the diagonal elements. The Jordan structure of a matrix A is unique up to a reordening of the Jordan blocks and is denoted by $J(A)$. In this context, $J(A) = J(B)$ summarizes the statement that both A and B have the same Jordan structure up to a reordering of the Jordan blocks.

Theorem 5 SVD of a rank deficient block Hankel matrix.
Let H be a $il \times jm$ block matrix of rank n with $n < i$ and $n < j$ and with a SVD : $H = USV^t$ where S is square diagonal $n \times n$. Then H is a block Hankel matrix if and only if the singular values are the solution to the set of linear equations:

$$[\overline{V} \odot \underline{U} - \underline{V} \odot \overline{U}]vecd(S) = 0$$

Proof: The theorem follows immediately from corollary 1 and the properties of the Khatri-Rao product. □

If the singular vectors of a matrix are known and if that matrix is to be of block Hankel structure, the singular values are the solution to a set of linear equations, of which the data are determined by the components of the singular vectors. In the next two theorems, it is investigated which block Hankel matrices could be constructed from two available matrices U $(il \times n)$ and V $(jm \times n)$ with shift structure and with shift matrices that have the same set of minimal system poles. The motivation arises from lemma 4 and theorem 5: If U and V contain the left and right singular vectors of a block Hankel matrix H, then the vector with singular values $vecd(S)$ belongs to the kernel of the matrix $[\overline{V} \odot \underline{U} - \underline{V} \odot \overline{U}]$. If now the orthonormal matrices U and V are modified into UP and VQ where P and Q are unitary, then, from lemma 4 the minimal system poles remain unchanged. In order to obtain a block Hankel matrix having singular vectors UP and VQ, according to theorem 5, the 'new' singular values contained in the vector $vecd(S_{PQ})$ will have to satisfy:

$$[(\overline{V}Q) \odot (\underline{U}P) - (\underline{V}Q) \odot (\overline{U}P)]vecd(S_{PQ}) = 0$$

or from the properties of the Kronecker and Khatri-Rao products:

$$[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}][Q \odot P]vecd(S_{PQ}) = 0$$

which is equivalent with:

$$[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]vec(PS_{PQ}Q^t) = 0$$

Hence, any vector $vec(X)$ of the kernel of the matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$ where X is a $n \times n$ matrix with SVD $X = P_x S_x Q_x$ will generate a block Hankel matrix UXV^t with SVD $(UP_x)S_x(VQ_x)^t$. Hence, the kernel of the matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$ plays a crucial role :

Theorem 6 *Let U $(il \times n)$ and V $(jm \times n)$ be orthonormal matrices with shift structure and with shift matrices T_u and T_v defined from $\underline{U}T_u = \overline{U}$ and $\underline{V}T_v = \overline{V}$, such that $J(T_u) = J(T_v) = J$. Then:*

$$cor[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}] = n$$

Proof: Let T_u and T_v have eigendecompositions:

$$T_u = X_u J X_u^{-1} \quad T_v = X_v J X_v^{-1}$$

Then with the properties of the Kronecker product:

$$\begin{aligned}
[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}] &= [(\underline{V}T_v) \otimes \underline{U} - \underline{V} \otimes (\underline{U}T_u)] \\
&= [\underline{V} \otimes \underline{U}][T_v \otimes I_n - I_n \otimes T_u] \\
&= [\underline{V} \otimes \underline{U}][(X_v J X_v^{-1}) \otimes I_n - I_n \otimes (X_u J X_u^{-1})] \\
&= [\underline{V} \otimes \underline{V}][X_v \otimes X_u][J \otimes I_n - I_n \otimes J][X_v \otimes X_u]^{-1}
\end{aligned}$$

From the partial realization condition the matrix $[\underline{V} \otimes \underline{U}]$ is non-singular as is the matrix, $[X_v \otimes X_u]$. It is easily verified that $cor[J \otimes I_n - I_n \otimes J] = n$. □

It will now be proved that every vector $vec(X)$, where X is a $n \times n$ matrix, in the kernel of the matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$ can be associated with a block Hankel matrix.

Theorem 7 *Let U $(il \times n, i > n)$ and V $(jm \times n, j > n)$ be orthonormal matrices with shift matrices T_u and T_v such that $J(T_u) = J(T_v) = J$. Let $vec(X)$ be a vector in the kernel of $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$:*

$$[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]vec(X) = 0$$

where X is a $n \times n$ matrix. Suppose $rank(X) = r$ and that X has an SVD $X = PSQ^t$ where S is square diagonal $r \times r$. Then the matrix $H = UXV^t$ is a block Hankel matrix of rank r, of which the r minimal system poles are a subset of the eigenvalues of J. An SVD of H reads $H = (UP)S(VQ)^t$.

Proof: Using the properties of the Kronecker and Khatri-Rao product:

$$\begin{aligned}
\left[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}\right] vec(X) &= 0 \\
\left[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}\right] (Q \odot P)vecd(S) &= 0 \\
\left[(\overline{V}Q) \odot (\underline{U}P) - (\underline{V}Q) \odot (\overline{U}P)\right] vecd(S) &= 0
\end{aligned}$$

From this it is obvious that:

$$\begin{aligned}
[(\overline{V}Q) \odot (\underline{U}P)]vecd(S) &= [(\underline{V}Q) \odot (\overline{U}P)]vecd(S) \\
(\underline{U}P)S(\overline{V}Q)^t &= (\overline{U}P)S(\underline{V}Q)^t
\end{aligned}$$

From theorem 1 it now follows that the matrix $H = (UP)S(VQ)^t$ is a block Hankel matrix. Moreover, if $rank(X) = r$, there exist $r \times r$ matrices T_{ur} and T_{vr} such that:

$$(\underline{U}P)T_{ur} = (\overline{U}P) = (\underline{U}T_u)P$$

and

$$(\underline{V}Q)T_{vr} = (\overline{V}Q) = (\underline{V}T_v)Q$$

Let T_{ur} and T_{vr} have the eigenvalue decompositions

$$T_{ur} = X_{ur} J_{ur} X_{ur}^{-1} \quad T_{vr} = X_{vr} J_{vr} X_{vr}^{-1}$$

where J_{ur} and J_{vr} are the Jordanform of T_{ur} and T_{vr}. Then:

$$(PX_{ur})J_{ur} = T_u(PX_{ur}) \quad (QX_{vr})J_{vr} = T_v(QX_{vr})$$

so that the diagonal elements of J_{ur} and J_{vr} are also eigenvalues of T_u and T_v. It then follows that $J_{ur} = J_{vr}$ □

Theorem 7 provides a parametrization of all block Hankel matrices with a given set of minimal system poles. This set of minimal system poles determines the column and row space of all block Hankel matrices with those minimal system poles. Once such a basis for this column and row space are known in the form of two matrices U $(il \times n)$ and $V(jm \times n)$, all vectors $vec(X)$ in the kernel of the matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$ result in a block Hankel matrix given by $H = UPS(VQ)^t$ where PSQ^t is the SVD of X. Clearly, the transformations $U \rightarrow UP$ and $V \rightarrow VQ$ correspond to a change of basis in the column- and rowspace and the singular values are to be adapted in order to preserve the block Hankel structure.

Theorem 7 can also be interpreted in terms of the zeros of the transferfunction $G(z) = C(zI - A)^{-1}B + D$. For single input single output, the denumerator of $G(z)$ only depends upon eigenvalues of the matrix A. The choices for other Hankel bases $U \rightarrow UP$ and $V \rightarrow VQ$ correspond to a change of the matrices B, C while the eigenvalues are fixed. Hence, the basis transformations correspond to assigning a new set of zeros to the transfer function $G(z)$. When X is a $n \times n$ matrix such that $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]vec(X) = 0$, then for $rank(X) = n$, no pole zero cancellations have occurred. This is the generic situation. However, if $rank(X) = r < n$, then the new assignment of zeros caused $n - r$ pole-zero cancellations.

7.3.2 An example

Consider the 4×4 system matrix A:

$$A = \begin{pmatrix} 0.9 & 1 & 0 & 0 \\ 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0.7 & -0.6 \\ 0 & 0 & 0.6 & 0.7 \end{pmatrix}$$

with a double eigenvalue 0.9 and a pair of complex conjugated eigenvalues $0.7 \pm 0.6j$. In order to parametrize all possible 8×6 Hankel matrices of rank 4 (or less),having this set of minimal system poles (or a subset of it), the following procedure can be applied:

- Generate a 8×4 orthonormal matrix U and a 6×4 orthonormal matrix V, both with shift structure. Hereto, chose a 'generator' $w^t = [1\ 1\ 1\ 1]$ (which has components along

all eigenvectors of A) and iterate to construct the 8×4 matrix W:

$$W = \begin{pmatrix} w^t \\ w^t A \\ w^t A^2 \\ w^t A^3 \\ w^t A^4 \\ w^t A^5 \\ w^t A^6 \\ w^t A^7 \end{pmatrix} = \begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.9000 & 1.9000 & 1.3000 & 0.1000 \\ 0.8100 & 2.6100 & 0.9700 & -0.7100 \\ 0.7290 & 3.1590 & 0.2530 & -1.0790 \\ 0.6561 & 3.5721 & -0.4703 & -0.9071 \\ 0.5905 & 3.8710 & -0.8735 & -0.3528 \\ 0.5314 & 4.0744 & -0.8231 & 0.2771 \\ 0.4783 & 4.1984 & -0.4099 & 0.6879 \end{pmatrix}$$

- The desired orthonormal matrices U and V can now easily be constructed by orthogonalising the columns of (parts of) the matrix W (e.g. with a QR factorization [8]). This is allowed from lemma 4, since such an orthonormalization corresponds simply to a change of basis with preservation of the column space. For V, take the first six rows of W and do a QR factorization. For U, do a QR factorization on the complete matrix W. U and V are now orthonormal matrices and it follows from lemma 4 that both U and V have the desired shift structure .

- Compute the 35×16 matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$. It can be verified numerically that the rank of this matrix is 12 (as is predicted by theorem 6). Let a basis of the kernel of this matrix be generated by the columns of the 16×4 matrix V_2

  $$[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]V_2 = 0$$

  Then from theorem 6 it follows that any linear combination of the columns of V_2 of the form $V_2 x = vec(X)$, where x is an arbitrary 4×1 vector, generates a vector $vec(X)$, such that X is a 4×4 matrix with the property that $H_X = UXV^t$ is a Hankel matrix. The kernel of the matrix $[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}]$ generates all Hankel matrices with the prespecified set of minimal system poles (or at least a subset of it, if pole-zero cancellations occur).

There is still some freedom left to pick those Hankel matrix that besides the prespecified poles have some other desirable properties e.g. prespecified zeros. This freedom will be used in a section 8.3.4. to make the Hankel matrix as orthogonal as possible.

7.3.3 Orthonormal matrices with shift structure.

In this section, we shall state and prove some nice properties of orthonormal matrices with shift structure. These are matrices of the form:

$$U = [\, g \;\; Tg \;\; T^2 g \;\; \ldots \;\; T^{j-1} g \,]$$

Here g is an $n \times 1$ vectors, T is square $n \times n$, and U is a $n \times j$ matrix, which is orthonormal:

$$UU^t = I_n$$

We shall restrict ourselves in this section (without loss of generality) to a shift structure generated by a vector g instead of a block column.

Observe that general matrices with shift structure become more and more ill conditioned for increasing j. The reason is that the iteration $T^k g$ converges under fairly general conditions, to the eigenvector of T corresponding to the eigenvalue of T with the largest modulus. Hence, the matrix with shift structure $[\ g\ Tg\ T^2g\ \ldots]$ will behave more and more as a rank one matrix. An orthonormal matrix with shift structure is however ideally conditioned. Hence, one could expect that the sensitivity properties of orthonormal matrices with shift structure are superior with respect to non-orthonormal matrices with similar shift structure. It will be demonstrated that the matrix T in an orthonormal matrix U with shift structure satisfies some remarkable properties.

Theorem 8 *Let U be an $n \times j$ orthonormal matrix with shift structure, generated by a $n \times 1$ vector g and a $n \times n$ matrix T, then:*

1. *$n-2$ singular values of T are equal to 1.*
2. *If $j = \infty$ and T stable, then $n-1$ singular values of T are equal to 1. The remaining singular value σ is the smallest and $g = \sqrt{(1-\sigma^2)}u_n$ where u_n is the corresponding left singular vector of T.*
3. *If $j = \infty$ and T stable, the singular values of T are the cosines of the canonical angles between:*

$$span_{row}(U) \quad and \quad span_{row}(U_1)$$

where $U_1 = [\ 0\ g\ gT\ gT^2\ \ldots\ ]$

Proof: 1. Multiply $UU^t = I_n$ to the left with T and to the right with T^t:

$$TT^t = I_n - gg^t + (T^j g)(T^j g)^t$$

Hence:

$$(TT^t - I_n) = -gg^t + (T^j g)(T^j g)^t$$

Assume that the SVD of $T = U_T \Sigma_T V_T^t$, then:

$$U_T(\Sigma_T^2 - I_n)U_T^t = -gg^t + (T^j g)(T^j g)^t$$

Because the right hand side is maximally of rank 2, at least $n-2$ singular values of T must be 1.

2. For a stable matrix T and with $j = \infty$, one finds that:

$$U_T(\Sigma_T^2 - I_n)U_T^t = -gg^t$$

Hence $n-1$ singular values of T are 1. Because gg^t is nonnegative definite, every singular value must be smaller than 1. Hence, the remaining singular value $\sigma \leq 1$. Moreover, it is easily derived that:

$$gg^t = (1-\sigma^2)u_n u_n^t$$

or:

$$g = \sqrt{(1-\sigma^2)}u_n$$

3. Observe that $T = UU_1^t$. This follows from $UU^t = I_n$ when multiplied on the left with T:

$$\begin{aligned} T &= [\,Tg\; T^2g\; \ldots\,] \begin{pmatrix} g^t \\ (Tg)^t \\ \cdots \\ \cdots \end{pmatrix} \\ &= [\,g\; Tg\; T^2g\; \ldots\,] \begin{pmatrix} 0 \\ g^t \\ (Tg)^t \\ \cdots \\ \cdots \end{pmatrix} \\ &= UU_1^t \end{aligned}$$

Observe that both U and U_1 are orthonormal matrices. Hence the singular values of the 'inner product' can be considered as the cosines of the canonical angles between the row space of U and U_1. □

7.3.4 Orthonormal rankdeficient Hankel matrices.

Orthonormal matrices with shift structure arise in the singular value decomposition of rank deficient (block) Hankel matrices. In this section, we shall go a little step further and state some new properties of scalar Hankel matrices that are themselves (partial) orthonormal. The qualification *partial* is to be understood in the sense that all singular values are either one or zero. Such Hankel matrices occur in two important forms in systems theory.

All Pass Transferfunction and Optimal Design.

In section 7.2., we have seen how the singular values of the (block) Hankel matrix with Markov parameters of a linear system allow to quantify degrees of controllability and observability via oriented energies in the state space. Assume that the minimal order of the system is n. Then, obviously, controllability and observability are jointly optimal if the restricted condition number $\sigma_1/\sigma_n = 1$, where σ_1 and σ_n are the largest and smallest non-zero singular value of the Hankel matrix. Optimality is to be interpreted in the sense of oriented energy: If all singular values of the Hankel matrix are equal, all possible states require equal energy to control and to observe. There are no 'difficult' regions in the state space that require a lot of energy to control or that contribute little energy to the output. Of course, there should be a kind of physical motivation in order to require such an optimality. However, we shall not go into detail but rather have a closer look at the mathematical problem. We shall restrict ourselves from now on to the scalar case only (single input - single output).

Obviously, the singular vales and the restricted condition number are influenced by the dimensions i and j of the Hankel matrix. These dimensions are equal to the time horizon for control (over j steps) and for observation (over i steps). If the system is stable, and both time horizons are infinite, the following result, due to Glover, allows to conclude that the corresponding double infinite Hankel matrix has restricted condition number one.

Theorem 9 All Pass Transfer Function.
Given a system triplet (A, b, c) for a continuous time single input single output system . If (A, b, c) is minimal, then there exist a real d such that:

1. $G(s)G^*(-s) = \sigma^2 I \quad s \in C$ *with* $G(s) = d + c(sI - A)^{-1}b$
2. *The non-zero Hankel singular values of $G(s)$ (which are the singular values of the double infinite Hankel matrix with the Markov parameters of the system) are all equal to 1.*

Proof: see [9]. □

$G(s)$ defines an all pass transfer function. Via the bilinear transformation $s = (z-1)/(z+1)$ one can derive a complete similar result for the discrete time case since it can be proven that the Hankel singular values are identical in the continuous and discrete time case [9]. Glover's result solves the design problem only for infinite time horizon. In an heuristic attempt to solve the problem, one could of course truncate the infinite Hankel matrix in order to obtain a finite Hankel matrix. However, the condition number can then be far from optimal as the following second order example suggest.

Example:
Consider the discrete time all pass transfer function:

$$G(z) = (0.98001z^2 - 1.9799z + 1)/(z^2 - 1.9799z + 0.98001)$$

with two very close real poles $p_1 = 0.99$ and $p_2 = 0.9899$. The following table gives the restricted condition number $K = \sigma_1/\sigma_2$ for several finite square truncations of the infinite Hankel matrix of Markov parameters.

k	K	k	K
4	7233.1	12	646.2
6	2977.5	14	455.4
8	1589.0	16	334.7
10	971.5	18	254.0
		20	197.7

Hence, for small time horizon k, the restricted condition number is far from the optimal 1. For increasing k, the restricted condition number converges to the optimal 1. The rate of convergence is however slow.

This means that it is meaningful to consider the problem:

Given a finite time horizon i for observation and time horizon j for control, and a desired set of minimal system poles J where J represents a certain Jordan structure. Compute a minimal system (A, b, c, d) with $J(A) = J$, such that the Hankel matrix H with Markov parameters $h_k = cA^{k-1}b$ has a minimal restricted condition number.

The problem is unsolved in its full generality. It is even not known whether the minimum restricted condition number 1 is achievable for all possible i and j. Glover's result demonstrates that it is always possible for double infinite time horizon.

However, the results obtained on the parametrization of Hankel matrices via the kernel of certain matrices with shift structure, allows to obtain more insight into the solution. The answer is to be found via the kernel of the matrix:

$$\left[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}\right]$$

where U is a $i \times n$ and V is a $j \times n$ matrix with appropriate shift structure. From theorem 6, it follows that the kernel is n-dimensional. Let the columns of the $n^2 \times n$ matrix V_2 span the kernel of this matrix.

$$\left[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}\right] V_2 = 0$$

Let $vec(X) = V_2 x$ be a linear combination of the columns of V_2 where X is a $n \times n$ matrix. The design problem is now reduced to the computation of all vectors x such that the condition number of X is 1. The obstacle for an easy derivation of the general solution is however the non-trivial relation between the linear combinations $V_2 x$ and the resulting matrix X.

For second order systems however, we have observed that the 4×2 kernel matrix V_2 has always the following structure:

$$V_2 = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \\ \pm\beta_1 & \pm\beta_2 \\ \gamma_1 & \gamma_2 \end{pmatrix}$$

Now it is easily verified that 2×2 orthonormal matrices can be parametrized as:

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \quad or \quad \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix}$$

where α and β are noth both 0. Hence any linear combination of the columns of V_2 that creates a vector $vec(X)$ such that X is a 2×2 matrix with condition number 1, will solve the problem.

Example:
Consider the impulse response of the system characterized by the transfer function $G(z)$ of the previous example.

$$\begin{aligned} h_k = 0.98001E+00 \quad & -3.9596E-02 \quad -3.8798E-02 \\ -3.8012E-02 \quad & -3.7238E-02 \quad -3.6476E-02 \\ -3.5725E-02 \quad & -3.4986E-02 \quad -3.4258E-02 \\ -3.3541E-02 \quad & -3.2835E-02 \quad -3.2140E-02 \ \ldots \end{aligned}$$

for $k = 0, 1, \ldots$. Construct with $h_1, \ldots, h_{11}$ a 6×6 Hankel matrix H and compute its singular value decomposition $H = USV^t$ where S is 2×2. H is of rank 2 and it can be verified that its restricted condition number equals 2977.5. A basis of the kernel of the 25×4 matrix $\left[\overline{V} \otimes \underline{U} - \underline{V} \otimes \overline{U}\right]$ is formed by the columns of:

$$V_2 = \begin{pmatrix} 9.7019E-04 & -1.0000E+00 \\ 7.0687E-01 & -6.7965E-04 \\ -7.0687E-01 & 6.7965E-04 \\ 2.5908E-02 & -3.6076E-04 \end{pmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \\ -\beta_1 & -\beta_2 \\ \gamma_1 & \gamma_2 \end{pmatrix}$$

The linear combination x of columns of V_2 such that $vec(X) = V_2 x$ where X is a 2×2 matrix with condition number one can now be computed from the condition:

$$x_1\alpha_1 + x_2\alpha_2 = x_1\gamma_1 + x_2\gamma_2$$

where x_1, x_2 are the components of x. For $x_1 = 1$ this yields $x_2 = -2.6888E-02$ and:

$$X = \begin{pmatrix} 2.5918E-02 & -7.0689E-01 \\ 7.0689E-01 & 2.5818E-02 \end{pmatrix}$$

which has condition number 1. The 6×6 matrix $H = UXV^t$ is now a Hankel matrix of rank two with restricted condition number 1 and with the same minimal system poles as the original Hankel matrix H. Its elements are

$$\begin{array}{lll} -3.5441E-01 & -2.7898E-01 & -2.0504E-01 \\ -1.3255E-01 & -6.1504E-02 & +8.1311E-03 \\ +7.6372E-02 & +1.4324E-01 & +2.0876E-01 \\ +2.7294E-01 & +3.3582E-01 & \end{array}$$

White Noise Hankel Matrices.

Assume that $h_k, k = 1, \ldots,$ are samples of a stationary white noise stochastic process and let H be the $i \times j$ Hankel matrix constructed from it.

Theorem 10 White noise and Hankel structure.
Let the $i \times j$ matrix H be a Hankel matrix containing samples of stationary white noise process with elementwise variance σ^2. Then:

$$lim_{j\to\infty} \frac{1}{j}(HH^t) = \sigma^2 I_i$$

Proof: The proof follows immediately from the observation that element (p, q) of $j^{-1}HH^t$ is nothing else but the $\mid p - q \mid$ value of the autocorrelation function of the process. □

Hence, in order for the oriented energy of a Hankel matrix to be *isotropic*, it needs to contain white noise! Observe that the asymptotic equality of all singular values of H implies that H will tend to be more and more orthonormal for increasing overdetermination j/i. Since orthonormality is equivalent with absence of linear relations, as was explained in chapter 5, a white noise signal is most remote from a linear dynamic system, because a linear system reveals itself via the rank deficiency of the Hankel matrix. This indicates that an extension of our results of chapter 6 (The uncertainty principle) towards dynamic system identification, could go via a splitting of a Hankel matrix of observations in a rank deficient one (of lowest possible rank) and an orthonormal one. At present, this problem remains however completely unsolved.

7.4 Structure exploiting factor analysis.

In chapter 4, we have encountered factor analysis applications, where the task was to obtain (good approximations to) certain subspaces. The criteria that are used are based upon oriented energy and signal-to-signal ratio distributions, expressing the hope that the 'real' system can be recovered from the disturbances (the noise) because this is of smaller energy content. In this section, it will be demonstrated that, under certain restrictive conditions, it is perfectly possible to recover *exactly* certain structured signals from disturbing, unstructured signals. The criterion is *not* based upon energy considerations, as in our signal-to-signal ratio framework of chapter 4. As a matter of fact, it exploits explicitly the shift structure. For this reason, we call it *structure exploiting factor analysis* and in our opinion its possibilities remain largely unexplored *terra incognita* in todays identification and modelling literature. The main difference between *subspace oriented energy based factor analysis* and *structure exploiting factor analysis* is that the latter allows for *exact* results, while the former one in the most favorable case only delivers good approximations. However, the possible disturbance models are more general in the energy based factor analysis approaches.

To be more specific, consider an $n \times jm$ matrix Z_1, with $n < j$ consisting of j $n \times m$ matrices Z_1^k:

$$Z_1 = [\ Z_1^1\ Z_1^2\ \dots\ Z_1^m\ ]$$

Assume that this matrix Z_1 has the following shift structure:

$$Z_1^{k+1} = AZ_1^k$$

where A is an $n \times n$ matrix. Its eigenvalues determine the *dynamics* of the shift. n will be called the order. This is written in shorthand notation as:

$$A(Z_1\ |) = (|\ Z_1)$$

where $(Z_1\ |)$ is the $n \times m(j-1)$ matrix, obtained from Z_1 by omission of the last block Z_1^m. $(|\ Z_1)$ is obtained by omission of Z_1^1. Consider also a $p \times mj$ matrix Z_2 with $p \leq j$, without shift structure. An $i \times jm$ matrix Z is now formed which consists of a static contribution from Z_1 and Z_2:

$$Z = T_1 Z_1 + T_2 Z_2$$

It is assumed that:

$$p \leq (i-n) \qquad 2i \leq j$$

T_1 is an $i \times n$ matrix of rank n and T_2 is a $i \times p$ matrix of rank $r \leq p$ such that

$$T_2^t T_1 = 0$$

The main result is stated in the following theorem:

Theorem 11 *Given the $i \times j$ matrix Z satisfying the conditions described above. Then the shift structure of the row space of Z_1, i.e. the order n and the shift dynamics (eigenvalues of A), can be computed from the matrices $(|\ Z)$ and $(Z\ |)$:*

- *The dimension of the subspace with shift structure is the number of zero singular values of:*

$$\begin{pmatrix} Z \mid \\ \mid Z \end{pmatrix}$$

Assume there are n zero singular values.

- *Assume that the matrix Q is an $n \times 2i$ matrix of rank n partitioned as $Q = (Q_1 \;\; Q_2)$, where Q_1 and Q_2 are $n \times i$ matrices, satisfying:*

$$(Q_1 \;\; Q_2) \begin{pmatrix} Z \mid \\ \mid Z \end{pmatrix} = 0$$

Then the eigenvalues that characterize the shift dynamics are the eigenvalues of the matrix:

$$-Q_1 Q_2^t (Q_2 Q_2^t)^{-1}$$

Proof: The proof is constructive. First, derive the following expressions:

$$(Z \mid) = T_1(Z_1 \mid) + T_2(Z_2 \mid) \qquad (\mid Z) = T_1(\mid Z_1) + T_2(\mid Z_2)$$

Because $T_2^t T_1 = 0$, there exists a $n \times i$ matrix P of rank n such that:

$$PT_2 = 0 \quad \textit{and} \quad \textit{rank}(PT_1) = n$$

Hence:

$$P(Z \mid) = PT_1(Z_1 \mid) \qquad P(\mid Z) = PT_1(\mid Z_1)$$

Because, PT_1 is non-singular:

$$(Z_1 \mid) = (PT_1)^{-1} P(Z \mid)$$

Since $A(Z_1 \mid) = (\mid Z_1)$, it follows that:

$$P(\mid Z) = [(PT_1)A(PT_1)^{-1}]P(Z \mid)$$

Therefore:

$$[\,((PT_1)A(PT_1)^{-1})P \;\; -P\,] \begin{pmatrix} (Z \mid) \\ (\mid Z) \end{pmatrix} = 0$$

First observe that the first matrix is determined up to a premultiplication with a non-singular $n \times n$ matrix. Moreover, from the structure of the first matrix and the fact that $\textit{rank}(P) = n$, it is straightforward to compute the matrix $(PT_1)A(PT_1)^{-1}$ which is similar with the matrix A, hence shares the same eigenvalues. A premultiplication with a non-singular $n \times n$ matrix will not affect the eigenvalues. □

The reason for the condition $2i \leq j$ can be clarified as follows: When it is not satisfied, there will exist vectors orthogonal to the column space of the matrix:

$$\begin{pmatrix} Z \mid \\ \mid Z \end{pmatrix}$$

even if none of the singular values is zero. Hence, a shift structure could exist but there would be no means to derive its dimension.

Example:
As a simple example, consider the matrix Z, constituted as:

$$Z = T_1 Z_1 + T_2 Z_2$$

where

$$T_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad Z_1 = (1\ \alpha\ \alpha^2\ \alpha^3\ \alpha^4)$$

and

$$T_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad Z_2 = (2\ 0\ -1\ 3\ 0)$$

Hence we have:

$$\left(\begin{array}{c|} Z \\ \hline \\ \end{array}\right) = \left(\begin{array}{cccc} 3 & \alpha & \alpha^2-1 & \alpha^3+3 \\ -1 & \alpha & \alpha^2+1 & \alpha^3-3 \\ \hline \alpha & \alpha^2-1 & \alpha^3+3 & \alpha^4 \\ \alpha & \alpha^2+1 & \alpha^3-3 & \alpha^4 \end{array}\right)$$

Hence, the matrix Q of theorem 11 becomes:

$$Q = (\alpha\ \alpha\ -1\ -1)$$

and the dynamics of the shift follow from:

$$-Q_1 Q_2^t (Q_2 Q_2^t)^{-1} = \alpha$$

7.5 Conclusions

In this chapter, it was shown how the shift structure of certain subspaces associated to block Hankel matrices is important from the point of view of realization. It was shown how some classical realization algorithms are an immediate consequence of this shift structure. Furthermore, the connection between oriented energy, the singular value decomposition and the qualifications of observability and controllability were explored. Using 'non-conventional' matrix calculus, it was shown how one can parametrize all rank deficient block Hankel matrices with a certain set of minimal system poles. Some special properties of orthonormal matrices with shift structure were derived, including orthonormal Hankel matrices. Finally, it was shown how the one-to-one relation between shift invariant subspaces and certain sets of eigenvalues, allows to obtain information concerning a system, if the shift structure is exploited. This leads to structure exploiting factor analysis.
All of these results will be applied intensively in chapter 8 on the identification of state space models from input-output data.

Bibliography

[1] Brewer J.W. *Kronecker Products and Matrix Calculus in System Theory.* IEEE Trans. on Circuits and Systems, Vol. CAS-25, no.9., sept. 1978.

[2] Bultheel A. *Numerieke Aspekten van de Adamjan-Arov-Krein Benadering.* Stelling voor het Doktoraat in de Wetenschappen, K.U.Leuven, September 1979.

[3] Damen A.A.H., Vandenhof P.M.J., Hajdasinski A.K. *The Page Matrix: An Excellent Tool for Noise Filtering of Markov Parameters, Order Testing and Realization.* Eindhoven University of Technology, EUT Report 82-E-127, August 1982.

[4] Damen A.A.H., Guidorzi R.P., Hajdasinski A.K. and Van den Hof P.M.J. *On multivariable partial realization.* Int. J. Control, Vol.41, No.3., pp.589-613, 1985.

[5] De Groen P., De Moor B. *The fit of a sum of exponentials to noisy data.* Journal of Computational and Applied Mathematics, 20, 1987, pp.175-187, North-Holland.

[6] De Moor B., Vandewalle J. *A numerically reliable algorithm for fitting a sum of exponentials or sinusoids to noisy data.* Proc. 3rd IFAC/IFIP International Symposium on Computer Aided Design in Control Engineering Systems. Lyngby, Copenhagen, Danmark, 1985.

[7] De Moor B., Vandewalle J. *Non-conventional matrix calculus in the analysis of rank deficient matrices of finite dimensions.* Systems and Control Letters, 9, North-Holland, 1987, pp. 401-410.

[8] Golub G.H., Van Loan C.F. *Matrix Computations.* North Oxford Academic, 1983.

[9] Glover K. *All Optimal Hankel Norm Approximations of Linear multivariable systems and their L^∞-error bounds.* Int. J. Control, 1984, Vol.39, no.6, 1115-1193.

[10] Graham A. *Kronecker products and matrix calculus with applications.* Ellis Horwood Series Mathematics and its Applications, John Wiley and Sons, New York.

[11] Ho B.L., Kalman R.E. *Effective construction of linear state variable models from input output data.* Proc. Third Allerton Conference, p.449-459, 1965.

[12] Iohvidov I.S. *Hankel and Toeplitz matrices and forms, algebraic theory.* (editor Gohberg), Birkhauser Boston, 1982.

[13] Jonckheere , Silverman . *A new set of invariants for linear systems , Application to Reduced Order Compensator Design.* IEEE Trans. Aut. Control, Vol. AC-28, n0.1, October 1983.

[14] Kailath T. *Linear Systems.* Prentice Hall, Information and System Sciences Series, 1980.

[15] Kabamba. *Balanced Gains and their Significance for L^2 Model Reduction.* IEEE Trans. Aut. Control, no.7, Vol. AC-30, July 1985.

[16] Kung S.Y. *A new Identification and Model Reduction Algorithm with Singular Value Decomposition.* Proc. 12th Asilomar Conf. Circuits Systems, Pacific Grove, California, November 1978, pp.705-714.

[17] Kung S.Y. *A Toeplitz Approximation Method and Some Applications.* Int. Symp. on Math. Theory of Networks and Systems, vol.4., Aug.5-7, 1981, Sancta Monica, California.

[18] Moore B.C. *Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction.* IEEE Trans. Aut. Control, Vol.AC-26, no.1, February 1981, p.17-32.

[19] Mullis C.T., Roberts R.A. *Synthesis of mimimum round-off noise fixed point digital filters.* IEEE Trans. Circuits and Systems, Vol.CAS-23, pp.555-562, 1976.

[20] Rao *A study of coefficient quantization errors in state space digital filters.* Proc. ICASSP 1985, March 26-29, 1985, Florida USA.

[21] Pernebo L., Silverman. *Model reduction via balanced state space representation.* IEEE Trans. an Automatic Control, Vol.AC-27, no.2, april 1982.

[22] Tether A.J. *Construction of minimal linear state variable models from finite input-output data.* IEEE Trans. Autom. Control, Vol.AC-15, No.4, pp.427-436 (1970).

[23] Verriest E.I. *Error analysis of linear recursions in floating point.* IEEE Conf. ASSP, Tampa, Florida, March 1985.

[24] Ramos J.A., Verriest E.I. *A unifying tool for comparing stochastic relaization algorithms and model reduction techniques.* Proc. of the 1984 Automatic Control Conference, San Diego, California, June 1984, pp.150-155.

[25] Zeiger H.P., Mc.Ewen A.J. *Approximate linear realization of given dimension via Ho's algorithm.* IEEE Trans. Automatic Control, AC-19, p153, 1974.



Cause-Effect Links

Chapter 8

Identification of state space models

8.1 Introduction

Identification is the process of constructing a mathematical model of a dynamical system from observations [34]. Mathematical models are typically used in science and applications to extract essential features from complicated evidence, to quantify implications, to gain a deeper understanding of complicated phenomena or to develop optimal control strategies. Mathematical models may be used in all those situations where experiments on the real system are too complicated (ecological systems), too dangerous (nuclear power plants), too expensive (loss of production), too time consuming (slow phenomena) or where the required hard knowledge of the system is lacking. Mathematical models may be able to uncover shortcomings and anomalies and can be used for monitoring and training via simulators. Prediction by a dynamical model is important in control system design, when this is more ambitious than the traditional trial-and-error tuned PID controller based on second order models. Especially since the demonstration of the solvability of optimal filtering and control problems, based upon quadratic optimization criteria, the need for dynamic models has increased. Central in a lot of approaches is the notion of a *state*, requiring appropriate state space models.

There is no need to emphasize that all these requirements, when fulfilled, will result in considerable *economical benefits.*

However, obtaining dynamic models from measurements on systems is not easy. A decade ago, the domain was described as *a bag of tricks within a fiddler's paradise* [14]. This statement is confirmed in [48] where it is said that *identification theory and time series analysis (...) is very much an area where some of the first principles still need to be sorted out.*

The reasons for this statement are the following. First of all, there is a wealth of model structures that may be studied or identified: parametric versus non-parametric (difference equations versus impulse response approaches), time domain versus frequency domain (transfer functions), state variable versus input-output, stochastic versus deterministic, ... There is not only an immense diversity in the models, the number of identification approaches and techniques may even be larger [6] [14] [31] [34], making a comparison very difficult, if not unfeasible. While a lot of these approaches are based upon Kolmogorov's probabilistic framework [31] [35], it is only recently that some alternative approaches have been proposed [24] [37] [48]. For instance, the approach by Willems is a careful rigorization and formalization of *approximate modelling of dynamical systems.*

During the past decade, a lot of attention was paid to the identification of *input-output models* [17] [31] [34]. This can be explained by the fact that a lot of these strategies can be analysed using well understood statistical techniques and that they allow for a fast, adaptive, recursive implementation, which in a lot of cases is based upon a generalized least squares estimation algorithm.
The identification of *state space models*, has received much less attention. Not only does this approach requires more background of linear algebra than the input-output framework, but also the computational needs are much heavier. It is however our belief, that obtaining state space models directly does posses certain advantages with respect to the input-output framework. This will now be explained.
The model of interest is the well known linear state space model for discrete time lumped systems of finite state space dimension,, with observed and unobserved inputs:

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k^1 u_k + B_k^2 v_k \\ y_k &= C_k x_k + D_k^1 u_k + D_k^2 v_k \end{aligned}$$

In this expressions, the observed inputs are the $m_1 \times 1$ vectors u_k, the unobserved inputs are the $m_2 \times 1$ vectors v_k, the $n \times 1$ vector x_k is the state vector while the $l \times 1$ vector is the output vector. An index k denotes the time instant. $A_k, B_k^1, B_k^2, C_k, D_k^1, D_k^2$ denote matrices of appropriate dimensions.

While differential equations have been analysed and studied since Newton's and Leibnitz's time, the vivid interest in state space models stems from the mid-fifties and some strikingly successfull examples of state estimation applications in the 1960's and the early seventies, including the moon landings of the Apollo project. The reasons for this are the following:

1. A lot of conceptual results are most easily described in terms of a state space model: stability, controllability, observability,.... Moreover, state space models are more general than the conventional input-output models that describe only the minimal part of the system. Observe that via the introduction of a state, only a first order matrix vector recursion is necessary in order to simulate the system. This property makes the model very well suited for use in a digital computing environment.

2. A whole bunch of geometrical theory is available in terms of state space models, allowing for decoupling, tracking, regulation etc... [48].

3. Complicated optimal filtering problems such as filters with pole-placement (Luenberger filter) or optimal filtering (Kalman filter), are readily solved in a state space formalism as well as optimal control problems (Pontryagin, the Hamilton-Jacobi equation, quadratic regulation). The main contribution of modern systems and control theory is to be found in the recognition of the fact that optimal control is to be performed in terms of state variable feedback, instead of output feedback, as was the case in the classical control theory. Moreover, the state space formalism allows for numerically reliable algorithms, such as square root Kalman filtering [44].

4. State space model identification promises to be *numerically robust*. The availibility of numerically reliable software for the computation of the *singular value decomposition* [16] among other algorithms, has given a big stimulus in identification approaches based

upon *numerical linear algebra.* In comparison with the input-output models for instance, there is no explicit need for (pseudo-canonical, overlapping) parametrizations which may lead to ill-conditioning [19] [47].

5. On the other hand, the computational requirements are higher. It may however be expected, that taking into account the development of VLSI the last 2 decades and the vivid interest in parallel processing techniques, this will not present a serious drawback in future. As a matter of fact, most of the algorithms and approaches that are discussed and proposed in this chapter, have been implemented on personal computers without significant problems.

The main question that will be discussed in this chapter is the following:

Given observations on the inputs u_k and the outputs y_k, determine a suitable *state space model for these data.*

Of course, all difficulties are hidden in the word *suitable.* In order to refine its meaning, let's have a closer look at the three ingredients of mathematical modelling, which are:

1. The data-measurements-observations
2. The set of models
3. The criterion

The data - measurements - observations: • Data acquisition is determined by the application. In a lot of industrial applications, it is relatively easy to measure variables while in the so called socio-economic applications, obtaining reliable measurements may be a serious problem. But even in an industrial environment, the data acquisition requires considerable know-how and experience. Useful practical suggestions can be found in [6] [8] [17].

- One has to determine which variables are to be measured, how long and with which sampling time. While a certain redundancy or dependency in the variables can be detected by the linear algebraic techiques to be presented furtheron, the choice of the sampling time is a critical parameter which has to be fixed a priori.
- It has to be determined which data preparation actions will be undertaken such as filtering, peak shaving etc.... Note that these initial operations on the data may have a considerable influence on the result and quality of the identification. Suggestions on the data preparation, can be found in [8].
- Another important problem, especially in our singular value based algorithms, is the problem of scaling. In a lot of practical identification experiments, the measurements may differ in several orders of magnitude, e.g. a pressure may be measured in hecto-pascal while a thickness is measured in millimeter. However, if both measurements have a considerable influence on the process, this will not be reflected appropriately in the singular values that will mainly be determined by those measurements that are characterized by the largest absolute values, independent of the unities. Hence an appropriate scaling may be nessary.

- It may be observed that the data themselves and the purpose of the model may suggest a certain identification method: if only the outputs are noisy, linear least squares algorithms are preferable. If the model is meant to be descriptive based upon noisy inputs and outputs, one could consider total linear least squares strategies. If a prediction error is to be minimized, canonical correlation like strategies are to be preferred.

The set of models that will be used in this work consists of the dimensional, linear, discrete time, lumped dynamical systems, that can be represented by the well known state space formulas: *linearity* is in most cases not a property of the physical system but a self-imposed limitation of the kind of mathematical operations that are allowed. Of course, the user's preference for simplicity is one of the main motivations. *Models are a matter of inspiration, not of deduction* [48]. However, the class of systems we consider in this work, are more general from the application point of view than would appear at first sight:

- While most real systems are infinitely dimensional, finite dimensional state space models may provide an *approximation*, which performs sufficiently well for the purpose of the model (simulation, prediction, control).
- The lumpedness originates from the fact that what one models is the transfer function between specific input-output measurements of one sensor.
- Non-linearities of the physical system are usually dealt with by allowing time variant linear models. The key idea, already exploited in Lyapunov's stability analysis from the beginning of this century, is that a non-linear system may be approximated by a linear one, for small excursions around a working point. However, the fact that most data cannot be described *exactly* by *linear models*, immediately implies that mathematical modelling is a matter of *approximation*, requiring certain criteria that allow to pick out appropriate models.

The criterion by which a model is chosen, depends certainly upon the application and the purpose itself of the model. If a good *description* of a certain behavior is desired, a descriptive strategy such as total linear least squares, may provide good models. If however a reliable *prediction* is needed, linear least squares might be more appropriate, as is for instance the case in adaptive control. In any case, the optimization criterion should reflect two desirable properties of a model:

- A desirable feature of a model is its simplicity: One tries to *explain* a certain behavior using as few equations and relations as possible. In linear models, complexity measures are provided by the (approximate) rank of certain matrices.
- However, the misfit, i.e. that part of the data which is not explained by the noise, should be *small*, in some appropriately defined sense. This misfit could be a predictive error, or the oriented energy of the residuals etc....

Obviously, these two characteristics are in some sense conflicting: Simple models will not be powerful in explaining data, while models that explain a lot (small misfit) will tend to be complicated.

These three ingredients constitute an identification method, which is seen to be essentially a *mapping from data to a model.* One should define rigorously what is precisely meant by

identification. Identification criteria are *complexity and misfit* [48]. Under the imposed restriction of linearity, modelling then reduces to finding the linear model with *minimal misfit* for a certain complexity, or the model with *minimal complexity* for a certain misfit. As an example, consider all possible SISO systems, be it linear or nonlinear, dynamic or static, time variant or time invariant, with an input sequence u_k , $k = 1, \ldots$ and a corresponding output sequence y_k , $k = 1, \ldots$. They are all described by the model $y_k = 0u_k$. Of course, the value of this model is absolutely *nihil.* It explains everything, in much the same way as Wronski's *Supreme Law* [11, p.58]. Obviously, a good model must forbid certain things to occur. Hence, the central theme in modelling is *falsification.* In our work, which is based upon linearity, falsification coincides with orthogonality as will be explained furtheron.

Besides the already emphasised characteristics, the most essential features of our modelling approach, to be contrasted with existing identification approaches, are:

Conceptual : The description of properties of systems and models in terms of certain *subspaces.* This originates not only in the conceptual background of the methods, but also in the computational advantages that are provided by the use of subspaces. The definition itself of a subspace, as a non-trivial part of an ambient space, requires the notion of linearity and as a *dualisation,* of *orthogonality* by which everything what does not belong to the subspace, is decided to belong to its orthogonal complement, which again is a subspace. Hence, subspaces are the linear algebraic translation of the notion of *falsification.* The model is what does belong to a certain subspace and what is not explained by it (the noise!) belongs to its orthogonal complement. Observer that the subspace approach and the relation between linearity and orthogonality constitute precisely the mathematical results obtained for the uncertainty principle of mathematical modelling (chapter 6), hence this connection immediately suggests a possible generalization towards dynamical systems.

Numerical-Computational : In general, algorithms involving subspace concepts and properties, are preferable because no a priori choice of a basis is fixed, as is the case with all kinds of ((pseudo-)canonical) parametrizations. Hence, subspace computations are essentially *reference frame free,* as is the case in the most important physical theories. Via the singular value decomposition, subspaces may be computed in a numerically robust way. The singular values allow to estimate the complexity of the model while at the same time, they provide quantitative information about the *conditioning* of the results, i.e. their sensitivity with respect to disturbances.

Mathematical characterization of causal dependency : An important issue is the *causal dependency* direction. A good identification method should be able to decide whether a system is autonomous or not. Or in more general terms, which variables correspond to inputs and which ones are to be considered as outputs. Hence, *causal dependency* should not be considered as an a priori imposed fact, but as a mathematical problem. However, in a lot of (industrial) applications, the inputs are fixed or assumed to be known. In this cases, some complications may occur that will be investigated.

Geometrical inspired algorithms : For noisy data, three conceptually different approaches will be derived each based on another geometrical observation: linear least squares is based upon a projection argument, total linear least squares is based on a corrective

(descriptive) insight, while canonical correlation is based upon the geometrical concept of angles between subspaces.

Identification of unstable systems : Contrary to e.g. certain stochastically motivated theories, the identification of unstable systems poses no real problem in our framework, apart of course from some natural restrictions posed by the computation equipment, such as numerical overflow etc.... Recall that one of the important reasons of (adaptive) control is the stabilization of an unstable plant. Hence it is necessary to be able to identify unstable systems.

The fuzzy character of singular values : In this chapter, a lot of properties and results are statements that are very much alike the *logical law of the excluded middle:* Either an event is A, or it is not A (exclusive or). A typical linear algebraic notion that obeys this law, is the notion of rank of a matrix. The singular values of a matrix allow for a more 'continuous', *fuzzy*, characterization of the notion of rank. In terms of singular values, it makes perfectly well sense to characterize a matrix to be almost rank-deficient or close to another one of lower rank (e.g. via the Eckart-Young theorem). The reader should keep this in mind when analysing our *excluded-middle-like* statements.

In this chapter, the reader will find both more conceptually oriented results on the one hand and more pragmatic algorithmic procedures on the other hand. The conceptual part, mainly contained in section 8.2., is largely inspired by the recent work of Willems [48]. However, we are convinced that we have succeeded in making also some nice contributions and suggestions within the new framework. The results and algorithms contained within sections 8.3, 8.4, 8.5 are biased towards the *classical* deductive modelling approach, although some of the results are not too far away from an inspiration point of view. Notwithstanding important breakthroughs recently [24] [48], in our opinion much research remains to be done concerning identification. From our point of view, it are not as much the algorithms or the solutions to mathematical problems that form the bottlenecks (these are *technical* questions), but it are a farreaching *rigorization* and *conceptualization* which are still lacking. Therefore the reader should be mild in his judice about the results of the *pragmatic* sections 8.2, 8.3, 8.4. We hope to be able in near future to replace them by a mathematical theory of identification (approximation), which will be inspired by [48], but will as well contain a lot of the geometrical insights that are obtained in these sections.

This chapter is organised as follows:

In section 8.2 , we derive a little known, though *fundamental, structured, input-output matrix equation* which relates a block Hankel matrix with inputs to a block Hankel matrix with outputs. The properties of this equation are studied in detail, including persistancy of excitation measures and a refined analysis of its rank properties. Attention is paid to the mathematical characterization of causal dependency and to the identification of systems with delays.

In section 8.3 , it is shown how, from a *deductive* point of view, *linear least squares estimation* is appropriate in those circumstances where only the output sequence is corrupted by additive noise. An heuristic approach is discussed. The idea is to exploit the geometry of the input-output equation in order to derive estimates of the system matrices, via the *shift-structure* of certain subspaces.

In section 8.4 , we discuss what could be considered as a heuristic *total linear least squares approach.* It is appropriate when all data are subject to noisy perturbations. The geometrical idea is here to exploit not the structure of the matrices of the input-output equation, but the structure of bases for their kernels. This results in a very interesting observation concerning the *consistency* of the method from a *deductive* point of view, in this sense that, asymptotically, the approaximated subspaces are estimated without error. The tools that are used are the lever and orthogonality theorem of chapter 5.

In section 8.5 , we exploit the notion of *Nerode equivalence*, in order to prove that the intersection of the row spaces of past and future input-output block Hankel matrices, delivers the state sequence. Several methods, based on diffferent criteria are proposed and compared, including canonical correlation analysis, and the solution via solving sets of linear equations. Attention is paid to a solution in the short space, which allows for important conceptual interpretation and computational efficiency.

In section 8.6 , we conclude this chapter with some illustrations of the efficiency and reliability of our algorithms in industrial applications.

In section 8.7 , the reader may finally find some rest when reading the conclusions.

8.2 Exact properties of exact state space models.

Fundamental insight in the properties of linear systems in terms of *exact* data, is a *conditio sine qua non* for a thorough understanding of any scheme that aims at identifying linear models from *noisy* data. Observe that in our inspirationist framework (chapter 5), the adjective 'exact' has a double meaning: on the one hand it implies that the data are 'noiseless' (not corrupted by additive (measurement) noise), and on the other hand it implies that they were generated by a 'true' linear system. It is the intention of this section to analyse in detail some less known input-output properties of multivariable, lumped, finite dimensional, linear, time invariant, discrete time systems, described by the equations:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k \end{aligned}$$

In these equations, k is the discrete time index, u_k is the $m \times 1$ vector of measured inputs, x_k is the $n \times 1$ state vector, y_k the $l \times 1$ vector of outputs.
Observe that this model is a simplified version of the state space model presented earlier. The motivation behind this simplification is the following:

- The only feasible manner to deal with *time variance* in a *black box modelling* environment, is to apply the classical principle of *quasi-stationarity.* It is assumed that in a certain window of finite length, the data can be approximated by a time-invariant model. Via a gliding window approach or a strategy based upon exponential weighting, one then approximates the time-variant behavior. Of course, all this requires a (non-trivial) appropriate choice of the sampling rate, with respect to the time-constants of the time-variance of the model.

- The non-observed inputs will be dealt with in an indirect way: Their oriented energy will be identified with the oriented energy of the misfit-residuals. The characterization

of the unobserved inputs via their oriented energy distribution is sufficient in most applications as e.g. minimum variance Gauss-Markov estimation, which in a recursive form reduces to the Kalman filter, or in optimal control to model and decouple the disturbances.

- In order to understand the difficulties involved with identification of the state space model, it is necessary to have considerable insight into the properties of this *pure* exact state space model.

8.2.1 A fundamental input-output matrix equation.

The equation that is derived in this section is a straightforward exercise in algebraic manipulation. Yet, it is seldom exploited in the identification community, despite the fact that it provides fundamental insights into the geometry of any identification approach.

Define the following matrices:

$$Y_h = \begin{pmatrix} y_k & y_{k+1} & y_{k+2} & \cdots & \cdots & y_{k+j-1} \\ y_{k+1} & y_{k+2} & y_{k+3} & \cdots & \cdots & y_{k+j} \\ y_{k+2} & y_{k+3} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{k+i-1} & y_{k+i} & y_{k+i+1} & \cdots & \cdots & y_{k+i+j-2} \end{pmatrix}$$

This is a $(li) \times j$ block Hankel matrix consisting of output vectors.

$$U_h = \begin{pmatrix} u_k & u_{k+1} & u_{k+2} & \cdots & \cdots & u_{k+j-1} \\ u_{k+1} & u_{k+2} & u_{k+3} & \cdots & \cdots & u_{k+j} \\ u_{k+2} & u_{k+3} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{k+i-1} & u_{k+i} & u_{k+i+1} & \cdots & \cdots & u_{k+i+j-2} \end{pmatrix}$$

This is a $(mi) \times j$ block Hankel matrix consisting of input vectors.

$$H_t = \begin{pmatrix} H_0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ H_1 & H_0 & 0 & \cdots & \cdots & \cdots & 0 \\ H_2 & H_1 & H_0 & \cdots & \cdots & \cdots & 0 \\ H_3 & H_2 & H_1 & H_0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ H_{i-1} & H_{i-2} & H_{i-3} & \cdots & \cdots & H_1 & H_0 \end{pmatrix}$$

This is a $(li) \times (mi)$ lower triangular block Toeplitz matrix, containing the Markov parameters of the system, defined by: $H_0 = D$ and $H_k = CA^{k-1}B$ for $k > 1$. These Markov parameters are easily seen to coincide with the impulse response matrices for the case of discrete time

systems.

$$\Gamma_i = \begin{pmatrix} C \\ CA \\ CA^2 \\ \dots \\ \dots \\ CA^{i-2} \\ CA^{i-1} \end{pmatrix}$$

This is a $(li) \times n$ matrix. Observe that for $i = n$ this matrix coincides with the observability matrix of the linear system. Finally, define the $n \times j$ matrix X which contains the state vector sequence:

$$X = (\, x_k \; x_{k+1} \; x_{k+2} \; \dots \; \dots \; x_{k+j-1} \,)$$

Then, for any pair of block indices i and j the following matrix input-output equation holds true:

Theorem 1 Input-output matrix equation

$$Y_h = \Gamma_i X + H_t U_h$$

Proof: Trivial □

Note the ressemblance of the input-output matrix equation with the output equation of the state space model! This, at first sight, trivial input-output matrix equation forms one of the fundamental basic insights of our identification approach that will be developed in this chapter. Another matrix that will be used is the following:

$$\Delta_j = (\, B \; AB \; A^2B \; \dots \; A^{j-1}B \,)$$

Observe that for $i = n$, this is the controllability matrix. It is also a straightforward exercise to show that $\Gamma_i \Delta_j$ is an $li \times mj$ block Hankel matrix.

$$H_h = \Gamma_i \Delta_j$$

Observe that the state vector sequence can be expressed in terms of the extended controllability matrix and an input block Toeplitz matrix:

$$[\, x_k \; x_{k+1} \; \dots \; x_{k+j-1} \,] \;=\; [\, x_k \; Ax_k \; \dots \; A^{j-2}x_k \,] + \Delta_{j-1} \begin{pmatrix} 0 & u_k & u_{k+1} & u_{k+2} & \dots & \dots & u_{k+j-2} \\ 0 & 0 & u_k & u_{k+1} & \dots & \dots & u_{k+j-3} \\ 0 & 0 & 0 & u_k & u_{k+1} & \dots & u_{k+j-4} \\ 0 & 0 & 0 & 0 & u_k & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & u_k \end{pmatrix}$$

The $(j-1)m \times j$ upper triangular block Toeplitz matrix appearing in this expression is denoted by U_t.
A nice observation is the fact that the effects caused by an initial state $x_k \neq 0$ can be treated as being caused by an additional input:

Lemma 1 Initial state as an additional input
For a system (A, B, C, D) with initial state x_0, it is always possible to define a system (A, B', C, D') with one additional input, having $x'_0 = 0$ where $B' = (B\ Ax_0)$ and $D' = (D\ Cx_0)$. The additional input sequence is a Dirac impulse.

Proof:

$$\begin{aligned}
x_{k+i} &= A^i x_0 + A^{i-1}Bu_0 + A^{i-2}Bu_1 + \ldots + Bu_{i-1} \\
&= A^{i-1}[\,B\ \ Ax_0\,]\begin{pmatrix} u_0 \\ 1 \end{pmatrix} + A^{i-2}[\,B\ \ Ax_0\,]\begin{pmatrix} u_1 \\ 0 \end{pmatrix} + \ldots + [\,B\ \ Ax_0\,]\begin{pmatrix} u_{i-1} \\ 0 \end{pmatrix} \\
y_0 &= Cx_0 + Du_0 = (D\ Cx_0)\begin{pmatrix} u_0 \\ 1 \end{pmatrix} \\
y_k &= Cx_k + Du_k = (D\ Cx_0)\begin{pmatrix} u_k \\ 0 \end{pmatrix}
\end{aligned}$$

□

This result implies that an initial state different from zero can always be replaced by an additional input sequence, consisting of a Dirac impulse and an extra column Ax_0 in the input matrix. Hence, from now on we shall consider only systems with *initial state zero* in the theoretical analysis of the input-output properties. If necessary, the input is extended with an additional Dirac impulse input. Combining all the previous equations, it is possible to write the output block Hankel matrix as a sum of 2 matrix terms:

$$Y_h = H_h U_t + H_t U_h$$

where Y_h is the $li \times j$ output block Hankel matrix, $H_h = \Gamma_i \Delta_{j-1}$ is a $li \times m(j-1)$ block Hankel matrix with the Markov parameters, U_t is the $m(j-1) \times j$ upper block triangular block Toeplitz matrix with the first $j-1$ input vectors, H_t is a $li \times mi$ lower triangular block Toeplitz matrix with Markov parameters while finally, U_h is a $mi \times j$ block Hankel matrix with inputs. Hence, the output block Hankel matrix is the sum of the product of a block Hankel times a block Toeplitz and the product of a block Toeplitz times a block Hankel matrix!
As a final observation, observe that one can rewrite this equation as follows:

$$\begin{aligned}
Y_h = &\begin{pmatrix}
H_{j-1} & H_{j-2} & \ldots & H_2 & H_1 & H_0 & 0 & \ldots & 0 \\
H_j & H_{j-1} & \ldots & H_3 & H_2 & H_1 & H_0 & \ldots & 0 \\
H_{j+1} & H_j & \ldots & H_4 & H_3 & H_2 & H_1 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
H_{i+j-2} & H_{i+j-3} & \ldots & \ldots & \ldots & H_{i-1} & \ldots & H_1 & H_0
\end{pmatrix} \\
+ &\begin{pmatrix}
0 & 0 & \ldots & \ldots & \ldots & 0 & & u_k \\
0 & 0 & \ldots & \ldots & \ldots & u_k & & u_{k+1} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots \\
0 & u_k & u_{k+1} & u_{k+2} & \ldots & \ldots & & u_{k+j-2} \\
u_k & u_{k+1} & u_{k+2} & \ldots & \ldots & \ldots & & u_{k+j-1} \\
u_{k+1} & u_{k+2} & \ldots & \ldots & \ldots & \ldots & & u_{k+j} \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots \\
u_{k+i-1} & u_{k+i} \ldots & \ldots & \ldots & \ldots & u_{k+i+j-2} & &
\end{pmatrix}
\end{aligned}$$

where the $l \times m$ matrices H_k are the Markov parameters. Hence, the output block Hankel matrix can be written as the product of a block Toeplitz matrix with the Markov parameters and a block Hankel matrix with the input vectors. In shorthand notation, this becomes:

$$Y_h = H_T U_H = [\ H_T^1\ \ H_T^2\ ] \begin{pmatrix} U_H^1 \\ U_H^2 \end{pmatrix}$$

with an obvious partioning of H_T and U_H where $H_T^2 = H_t$ and $U_H^2 = U_h$.

8.2.2 Some persistancy of excitation results

The previous input-state-output equations allow to obtain a geometrical measure for the *persistancy of excitation* of an input sequence as the condition number of an upper triangular block Toeplitz matrix consisting of input vectors. However, first we need some definitions.

Definition 1 *Consider a linear system with the state vector sequence matrix X. Then:*

1. *The controllable subspace = $span_{col}(\Delta_n)$.*
2. *The observable subspace = $span_{row}(\Gamma_n)$.*
3. *The excited subspace = $span_{col}(X)$*

The dimensions of the controllable and observable subspaces will be denoted by n_c, respectively n_o. The dimension of the excited subspace is denoted by n_e. Note that both observable and controllable subspace only depend on the specific system, while the excited subspace also depends on the applied input sequence and/or the initial state. The excited subspace can be divided in a controllable part and an uncontrollable part. The controllable part originated in an applied input sequence while the uncontrollable part could only be caused by some non-zero initial state. The dimension of the intersection of the excited subspace with the controllable subspace is denoted by n_{ce}. It is called the *controllable excited dimension.* Obviously, $n_{ce} \leq min(n_e, n_c)$. In the same way, the excited subspace can be divided into an observable and an unobservable part. The dimension of the intersection of the excited subspace with the observability subspace will be called the *observable excited dimension.* It will be denoted by n_{oe}. Obviously, $n_{oe} \leq min(n_e, n_o)$.

Definition 2 *An input sequence u_p, $p = k, \ldots, j$ is persistantly exciting, iff $n_e = n_c$.*

Again, taking into account the *fuzzy* character of the singular values with respect to the notion of rank, one could also introduce more quantitative definitions, describing in a more quantitative way the persistancy of excitation in terms of singular values. However, we shall leave this exercise to the interested reader.

Consider with an initial state $x_k = 0$, the equality:

$$[\ x_{k+1}\ x_{k+2}\ \ldots\ x_{k+j-1}] \ =$$

$$[\ B\ AB\ \ldots\ A^{k-1}B\ ] \begin{pmatrix} u_k & u_{k+1} & u_{k+2} & \ldots & \ldots & u_{k+j-2} \\ 0 & u_k & u_{k+1} & \ldots & \ldots & u_{k+j-3} \\ 0 & 0 & u_k & \ldots & \ldots & u_{k+j-4} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \ldots & u_k \end{pmatrix}$$

or in shorthand notation:

$$X = \Delta_k U_t$$

Here U_t is obviously a $mk \times j$ upper triangular block Toeplitz matrix. Making use of the singular value inequalities for the product of two matrices, one easily finds for $k \geq n$:

$$K(X) = \frac{\sigma_1(X)}{\sigma_n(X)} \leq \frac{\sigma_1(\Delta_k)}{\sigma_n(\Delta_k)} \frac{\sigma_1(U_t)}{\sigma_n(U_t)} \leq K(\Delta_k)K(U_t)$$

where $K(.)$ denotes the condition number of a matrix. Note that the first inequality is in terms of the first and the n-th singular value of each matrix! The condition number of the matrix X can be taken as a measure for the oriented energy of the state vector sequence in the n-dimensional state space. If this condition number equals 1, all directions in the state space are equally well present. If however it tends to infinity, some of the directions have not been excited very well. The state vector sequence condition number is bounded by the product of 2 factors: The first one represents a quantitative measure of controllability, as explained in chapter 7. The second one can be derived from the input sequence. Input sequences that are such that the ratio $\sigma_1(U_t)/\sigma_n(U_t)$ is small, tend to force an isotropic oriented energy distribution of the state sequence. For instance, if the system has only one input ($m = 1$), an input signal which is perfect with respect to this criterion, is a Dirac impulse. Note that the second bound in terms of the condition numbers of both matrices, is much more conservative than the first bound: For instance, in the case where $u_k = 0$, the condition number $K(U_t)$ is infinite!

8.2.3 The main theorem.

In this section, we shall prove a very important theorem which is new but on which are based all the subsequent insights in the identification approach that will be developed furtheron. In essence, the theorem will *generically* allow to estimate the complexity of the model and also will lead to the identification of the model matrices. First, let's state the following lemma, which is a slight completion of a well known geometrical matrix property [15]: .

Lemma 2 The Generalized Sylvester Theorem
Let $C = AB$ where A is a $m \times n$, B is a $n \times p$ and C is a $m \times p$ matrix, of rank r_A, r_B, r_C respectively. Then:

$$r_C = min(r_A, r_B) - dim[span_{row}(A) \bigcap span_{col}^{\perp}(B)]$$

Proof : [15] □

Whenever $r_C < min(r_A, r_B)$, we shall say that *rank cancellation* has occurred in the multiplication of A and B.

Theorem 2 *Assume that the linear system is observable and that the dimension of the observable subspace equals n_o. Let the block dimensions i and j in the fundamental input-output equation be such that $i \geq n_o$ and $j \geq (mi + n_o)$. Denote the observable excited order by n_{oe}. Then:*

1. *The state sequence is given by:*

$$X = \Gamma_i^+ (I_{li} \quad - H_t) \begin{pmatrix} Y_h \\ U_h \end{pmatrix} \tag{8.1}$$

where $\Gamma_i^+ = (\Gamma_i^t \Gamma_i)^{-1} \Gamma_i^t$

2. *The rank of the concatenation of the input- and output- block Hankel matrices is given by:*

$$rank \begin{pmatrix} Y_h \\ U_h \end{pmatrix} = n_{oe} + rank(U_h) - dim[span_{row}(X) \bigcap span_{row}(U_h)] \tag{8.2}$$

The first part of the theorem states that the state vector sequence can be written in terms of linear combinations of the rows of the output and input block Hankel matrices. The second part of the theorem reveals how the observed excited order is related to the rank of the matrix which is the concatenation of input and output block Hankel matrices.

Proof: The first part follows immediately from the observability of the system, which implies that Γ_i is of full column rank, hence $(\Gamma_i^t \Gamma_i)^{-1}$ exists. In order to prove the second part, observe that the requirement for j is such that there exists an $j \times (j - mi)$ matrix $U_h^\perp$, which is at least of rank n_0, such that $U_h U_h^\perp = 0$. Right multiplication of the input-output equation with this matrix, results in:

$$Y_h U_h^\perp = \Gamma_i X U_h^\perp$$

Since Γ_i has full column rank:

$$rank(Y_h U_h^\perp) = rank(X U_h^\perp)$$

If the row spaces of X and U_h do not intersect, then $rank(X U_h^\perp) = rank(X) = n_{oe}$. If they intersect, then it is easy to see that $rank(X U_h^\perp) = rank(X) - dim[span_{row}(X) \bigcap span_{row}(U_h)]$. So in general, we have from lemma 2 that:

$$rank(Y_h U_h^\perp) = n_{oe} - dim[span_{row}(X) \bigcap span_{row}(U_h)]$$

which is precisely the dimensions of the subspace of the row space of Y_h which is not contained in the row space of U_h. This proves the theorem. □

The theorem has deep conceptual consequences. These will now be discussed in detail, and whenever possible, illustrated with some clarifying examples. The rank property is discussed in the next subsection, while attention is paid to the intersection property in section 8.5.

8.2.4 The analysis of the rank property.

The rank property stated in theorem 2, will allow us to determine the 'complexity' of the dynamic linear system that will be fitted through the observations. While results on rank properties of Hankel matrices containing the Markov parameters of a finite dimensional linear system go back to Kronecker [15], results on the rank of matrices containing input-output pairs on lumped linear systems are sparse. In recent literature, only [18], [29], [48] contain some useful observations. In what now follows, we provide a detailed discussion of the second

part of theorem 2.

As stated in theorem 2, the rank of the concatenation of Y_h and U_h equals the sum of three terms:

1. The observable excited dimensions n_{oe}.
2. The rank of the input block Hankel matrix
3. The phenomenon of rank cancellation

The observable excited dimension n_{oe}.

The system part corresponding to the observable excited subspace is the only part of the system that can be modelled. First, somewhat trivial, only the observable part of the state space can be observed by definition. Hence, only modes of operation that correspond to this subspace can be modelled. On the other hand, it is already intuitively clear that only those modes that have been excited, can be identified. What is not present in the data, can obviously not be extracted from it. Also with respect to control applications, this property is completely natural. In a lot of control applications, based upon the *certainty equivalence principle*, the control action essentially reduces to a gradual elimination of the dynamics of the system, as measured by the oriented energy distribution of the state (see chapter 7). It suffices to model only those modes that are not yet sufficiently well controlled. After a certain settling time, the controller will have removed all dynamics, that originated via the input. Observe however that at this moment, not necessarily $n_{oe} = 0$. Indeed, other modes might still be present that belong to the observable part of the state space that is not controllable. Observe that our approach differs fundamentally from other approaches described in literature. The reason is that we take it for granted, that **the ultimate goal or consequence of a control action, is the reduction of the dynamics of the system.** However, the algorithms that are used in existing identification algorithms will show numerically instabilities, when the input signals in the system have not been 'persistently exciting'. As an example, consider the widely spread recursive least squares techniques in parametric identification approaches used in an adaptive control environment. When some of the dynamics disappear, due to the succesful operation of the controller, *collinearities* will start to appear in the least squares matrices. The results delivered by the algorithm will start to behave very wild. This will produce on its turn, wild controller actions such that some modes are again excited and the story can start from the beginning. Of course, such a mode of controller operation is quite absurd. Several remedies are applicable:

1. Apply some regularization techniques. In the context of least squares, this reduces to adding to the Grammian of the normal equations, a multiple kI_n of the identity matrix, where k is a regularization parameter which makes the Grammian non-singular, so that it is invertible. However, by regularization one could also mean some more sophisticated ideas: For instance, in numerical linear algebra it is a good idea to restric the rank of certain matrices explicitly and aplly the algorithm with reduced rank matrices. As an example, consider the normal rank of a matrix pencil with a Jordan block. For arbitrary perturbations, the normal rank will increase. However, for restricted rank perturbations, the normal rank may be preserved. Also total least squares can be interpreted in a similar regularization context [43]. In general however, regularization

other effects that are not desirable such as biasedness though it may have advantagesd such as limiting the variance of the estimate etc...

2. Eliminate the cause of the ill-conditioning. In most cases, the bad behavior is caused by the fact that the inputs to the system, which are the outputs of the controller, become very smooth in time such that collinearities start to appear. The inputs are no longer persistently exciting. What is then advocated in the adaptive control literature is the introduction of a exciting additional input signal, independent of the controller, that excites, with small energy, the states, such that the parameters can be identified in a numerical reliable way. Of course, this is a disappointing philosphy. Suppose that one was able to design a really superior controller, then its power would be considerably reduced by this nasty little disturbance signal, which acts completely against the ultimate intention of the controller.

3. The only theoretically defendable approach that remains, consists of reducing the number of parameters as the complexity of the system decreases by the succesful controller action. Almost surely, this will require rank tests, for which the singular value decomposition provides the most reliable machinery. This is precisely what we achieve in our identification approach.

4. Observe that the order of dynamic phenomena, that can be modelled as the output of linear autonomous systems, can be estimated. Examples are:

   - Drifting signals
   - Sinusoidal Interference (e.g. 50 - 60 Hz)
   - Seaonal trends (econometrics)
   - Daily trends (day - night)
   - etc...

   The corresponding complexities are perfectly well identifiable from the singular values.

The rank of the input block Hankel matrix.

In an *open loop identification* set-up, (the precise mathematical reason will be explained in the sequel), the input block Hankel matrix will be of full row rank if the applied input sequence is rich enough in spectral content. Indeed, it is only for some very specific 'structured' signals, that the input block Hankel matrix may become rank deficient (see chapter 7). If for instance the input signal consists of a sum of p (damped) sinusoids, the rank of U_h will be $2p$ if $2p < mi < j$. However, as will be discussed furtheron, such signals are not to be considered as inputs *stricto sensu*. In fact, they are outputs of an autonomous linear system. In this way, they contradict our intuitive notion of an input as being a variable which is not determined by other variables of the same system, i.e. an independent variable (independent from other variables in the same system), which provides the interconnection of the outer world with the system. As a matter of fact, an input is really an input, if it has the privileged characteristic of *causing* other variables. Input variables are variables that cannot be explained by other variables in the same system, in contrast with states and outputs. Both states and outputs are caused by other states and inputs. As an example, consider the case of state variable feedback, which will also be considered furtheron. Despite the fact that one may keep measuring the

signals that were considered as inputs before the control loop is closed, these signals loose there significance as an input and as will be shown furtheron, they become essentially output variables. There is however another reason to prefer full row rank block Hankel matrices. Observe that the Grammian matrix $U_h U_h^t$ provides the finite sample statistical estimate of the first i values of the autocorrelation function of the input signal, if this was stationary. If the input is *white noise*, the autocorrelation function is a Dirac impulse and the finite block Hankel matrix will tend to be orthonormal for increasing overdetermination j/i. This confirms the idea of a white noise input signal being a 'good' identification signal. As a matter of fact, white noise can be considered as the absence of any linear cause opposite to the 'structured' signals that are outputs of an autonomous linear system. Obviously, the continuous transition from one extreme to the other is revealed by the singular values of the input block Hankel matrix U_h.

$\dim(span_{row}(X) \bigcap span_{row}(U_h))$ or the phenomenon of rank cancellation

This term can only differ from zero if the so called phenomenon of *rank cancellation* has occurred. This can happen in two situations: The first one is due to the fact that the input sequence is in fact a signal with structure. The occurence of the second case is rare and is due to mathematical coincidence.

Case 1: Dependent inputs.
The inputs themselves are outputs of an autonomous linear system, described by matrices A' and C' and states x'_k:

$$x'_{k+1} = A'x'_k \qquad u_k = C'x'_k$$

If the the excited observable dimension of this system equals n'_{oe}, then $rank(U_h) = n'_{oe}$ if $n'_{oe} < mi < j$. This follows directly from:

$$U_h = \begin{pmatrix} C' \\ C'A' \\ \cdots \\ C'A'^{i-1} \end{pmatrix} [\, x'_k \; x'_{k+1} \; \cdots \; x'_{k+j-1} \,]$$

Another consequence is that the rows of the input block Hankel matrix can now be written as linear combinations of the state sequence that has caused these rows. If this state sequence is contained in the matrix X', then obviously:

$$dim[span_{row}(X') \bigcap span_{row}(U_h)] = rank(U_h)$$

As an example, consider time invariant state variable feedback where the input is determined as $u_k = -Fx_k$. Then, when one keeps measuring the inputs, the system equations can be written as:

$$\begin{aligned} x_{k+1} &= (A - BF)x_k \\ \begin{pmatrix} y_k \\ u_k \end{pmatrix} &= \begin{pmatrix} C - DF \\ -F \end{pmatrix} x_k \end{aligned}$$

This illustrates clearly that the so-called 'inputs' are no longer free variables, but dependent on other variables in the system. The preceeding statement on state feedback, may be generalized as follows:

Lemma 3 *If there exists an $li \times n$ matrix L of rank n such that $U_h = LX$, then the inputs are determined by state variable feedback.*

Proof : Obviously U_h is a rank deficient block Hankel matrix. Hence, from chapter 7, the matrix L must have shift structure, which reveals the state variable feedback. □

Case 2: Coincidental rank cancellation The preceding case occurs if the input block Hankel matrix is not of full row rank. It is however feasible to construct examples of systems and input sequences with block Hankel matrix of full row rank, where rank cancellation occurs between the state matrix and the input block Hankel matrix. The mathematical analysis is highly complicated because the state is not (necessarily) independent of the input. If this were the case, then the proof of the *non-genericity* of rank cancellation would be straigtforward since the probability of intersection of the two row spaces of X and U_h decreases with increasing overdetermination j/i. However, we shall now prove the following remarkable theorem:

Theorem 3 *Assume that $r_{U_H} = rank(U_H)$ where U_H is the block Hankel matrix of input vectors as in section 8.2.1.with as always $mi < j$. Let the matrix P be such that:*

$$PU_H = [\ P_1\ \ P_2\ ] \begin{pmatrix} U_H^1 \\ U_H^2 \end{pmatrix} = 0$$

P is a $(m(j-1) + mi - r_{U_H}) \times (m(j-1) + mi)$ matrix of full row rank. P_1 is a $(m(j-1) + mi - r_{U_H}) \times (m(j-1))$ matrix. Then a necessary condition for rank cancellation to occur is that there exists a shift invariant subspace in $span_{row}(P1)$. The possible state sequence is then given by $X = -P_2 U_h$.

Proof: First note that lemma 1 allows us to assume that the initial state $x_k = 0$. Then the state vector sequence can be written as:

$$X = [\ x_k\ x_{k+1}\ \ldots\ x_{k+j-1}\ ] = \Delta_{j-1}^r U_H^1$$

Here $\Delta_{j-1}^r = [\ A^{k+j-2}B\ \ldots\ AB\ B\ ]$ is the reversed extended controllability matrix and U_H^1 is defined in section 8.2.1. Rank cancellation will occur if there exists a matrix Q_2 such that:

$$X = \Delta_{j-1}^r U_H^1 = Q_2 U_H^2$$

where $U_H^2 = U_h$. Or written in another form:

$$[\ \Delta_{j-1}^r\ \ Q_2\ ] U_H = 0$$

This proves the existence of a shift invariant subspace in the row space of P_1 as defined in the theorem. It can be computed with the aid of the structure exploiting factor analysis theorem 11 of chapter 7. □

In addition to theorem 3, let's mention the following remarks:

1. The matrices A' and B' that model the relation between the cancelled states and the input sequence, can be obtained from the shift structure of the *shift invariant subspace* of

maximal dimension, contained in $span_{row}(P_1)$, as demonstrated in theorem 11 chapter 7. Hence, we have proved that, associated to a certain input sequences and to prespecified block dimensions i and j, there may exists certain state sequences with corresponding system matrices A' and B' that are completely contained in the row space of the block input Hankel matrix.

2. Observe that the theorem only provides a necessary condition. It is possible that there exist matrices A' and B' and a candidate state sequence which could be cancelled by the row space of the input block Hankel matrix. However, rank cancellation also depends upon the actual system matrices A and B. Hence, it may be possible that there is no rank cancellation although the necessary condition is satisfied. The question remains how to detect a priori if rank cancellation has really occured. The simple answer is provided in the following result:

   Corollary 1 Detection of rank cancellation
   That rank cancellation has occurred, can be verified from the following procedure:

   (a) Compute the candidate state sequences for rank cancellation based upon theorem 3. If there are none, no rank cancellation is possible with this input sequence. If there are, proceed.

   (b) Compute the intersection of the row space of the candidate state sequence with the row space of the output block Hankel matrix Y_h. If there is a non-trivial intersection of dimension k, rank cancellation has occurred for k dimensions.

   Proof: The proof follows from a combination of theorem 3 and the fundamental input output provided in theorem 1. □

   Of course, a more pragmatic remedy is to construct a new block Hankel matrix with the inputs with other block dimensions i and j and see if the rank changes. Since rank cancellation depends upon the block dimensions i and j of the input block Hankel matrix, it is highly impropable that the phenomenon would persist over different choices of i and j. Observe however, that theoretically, this is not impossible because as was demonstrated under case 1, rank cancellation always occurs if the inputs (or part of it) are no *free variables*. Hence, if in a black box identification approach, one accidentally considers a dependent variable as an independent one, rank cancellation will occur for all block dimensions i and j. Hence, what is needed is a mathematical *causal dependency test*. This will be the subject of subsection 9.2.6.

3. Despite the fact that rank cancellation is also system dependent, it is now shown how it is always possible to construct an input sequence, which guarantees that no rank cancellation can occur, independent of the system. It suffices to avoid that a shift invariant subspace may exist in $span_{row}(P_1)$.

   Corollary 2 *For a zero state initial system with m inputs, there always exists at least one input sequence such that rank cancellation cannot occur.*

Proof: The proof is constructive. Take as an input sequence a Dirac impulse on all inputs at once.

$$[\, u_k \; u_{k+1} \; \ldots \; u_{k+j-1} \,] = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & \ldots & 0 \end{pmatrix}$$

Define the $(m-1) \times m$ matrix:

$$E_m = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & \ldots & \ldots & -1 \end{pmatrix}$$

Defining the $m \times 1$ vector $e = [\, 1 \; \ldots \; 1 \,]^t$, the input block Hankel matrix U_H is given by:

$$U_H = \begin{pmatrix} 0 & \ldots & \ldots & 0 & e \\ 0 & \ldots & \ldots & e & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & e & 0 & \ldots & 0 \\ e & 0 & \ldots & \ldots & 0 \\ 0 & \ldots & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots & 0 \end{pmatrix}$$

Obviously, its orthogonal column complement is generated by the rows of the $((m-1)j + m(i-1)) \times (m(j-1) + mi)$ matrix:

$$\begin{pmatrix} E_m & 0 & \ldots & 0 & 0 \\ 0 & E_m & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & E_m & \\ 0 & 0 & 0 & 0 & I_{m(i-1)} \end{pmatrix}$$

The matrix P_1 of the theorem is the upper $((m-1)(j-1)) \times (m(j-1))$ part of this matrix. However, observe that $P_1 U_H^1 = 0$. Hence, even if there is a shift structure (and there is one!), the corresponding state sequence is identically zero! □

Observe however that this input sequence is by no means the only possible one to avoid the occurrence of rank cancellation. Examples will be presented furtheron.

4. Finally, from the detailed analysis, it may be concluded that, for randomly generated input signals, and for sufficiently large overdetermination j/i, it is very unlikely that rank cancellation will occur in an identification experiment. If there is however any doubt we have provided all the tools to verify the possibility. Moroever, note that in a stochastic framework, the matter of rank cancellation would be passed off easily by requiring or assuming *ab initio* that $E(x_k u_k^t) = 0$, $E(x_k u_{k+1}^t) = 0$, ..., where $E(.)$ is the expectation operator. This assumption may be found in almost every paper treating a state space problem from a stochastic point of view.

8.2.5 Some small examples

In this section, it is our intention to clarify the detailed theoretical analysis of the rank property by some examples.

Example 1:
This example illustrates that the provided conditions for rank cancellation are only necessary. Hereto consider the system:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad C = [\,1 \;\; 1\,] \quad D = 0$$

with zero initial state. The input sequence and corresponding state and output are given in the following table:

k	0	1	2	3	4	5	6
u(k)	1.0000	-.2360	0.3820	-.6180	1.0000	-1.618	2.6180
x(k)	0.0000	1.0000	-.2360	0.3820	-.6180	1.0000	2.6180
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
y(k)	0.0000	1.0000	-.2360	0.3820	-.6180	1.0000	-.1618

Observe that the second state is not controllable. The excited observable dimension equals 1. Choose the parameters $i = 3$ and $j = 5$. The 7×5 matrix U_H is of rank 5. A suitable 2×4 matrix P_1 as in theorem 3, can be computed, e.g. with singular value decomposition.

$$P_1 = \begin{pmatrix} 0.3401 & -.2461 & 0.1718 & -.1289 \\ -.5503 & 0.3982 & -.2882 & 0.2085 \end{pmatrix}$$

It may be verified that this matrix has a shift structure of the form $P_1 = [\, A'^3b' \;\; A'^2b' \;\; A'b' \;\; b' \,]$ where:

$$A' = \begin{pmatrix} 0.1327 & 0.9361 \\ 0.7852 & -.8967 \end{pmatrix} \quad b' = \begin{pmatrix} -.1289 \\ 0.2085 \end{pmatrix}$$

The eigenvalues of A' are 0.6180 and -1.382. The state sequence X' that is generated by this system when the given inputs are applied, belongs by construction to the row space of the 3×5 Hankel matrix U_h.

$$X' = \begin{pmatrix} 0.0000 & -.1289 & 0.2085 & -.3374 & 0.5459 \\ 0.0000 & 2.0849 & -.3374 & 0.5459 & -.8832 \end{pmatrix}$$

It can be verified that the real state sequence cannot be written as a linear combination of this cancellable state sequence. Hence, rank cancellation does not occur for this system.

Observe that our technique provides at once a model (the matrices A' and b') for all possible systems such that rank cancellation is possible with this input sequences.

Example 2:
Consider the input sequence:

$$u(.) = [\,1\;2\;-3\;1\;-1\;-1\;2\;5\,]$$

and choose as block dimensions $i = 3$ and $j = 6$. Hence the block Hankel matrix U_H of theorem 3 is a 8×6 matrix. This matrix is of rank 6. Hence, the orthogonal complement of its column space is of dimension 2. However, the necessary condition for a shift structure .... is not satisfied. Hence, with this input, rank cancellation will be impossible, independent of the particular system! This example hence illustrates that there may exist other input sequences than the one constructed in corollary 2, such that rank cancellation is impossible.

Example 3:

Consider the same matrices A, B, C, D and the same input sequence as in example 1, but now take as initial state $x_0 = [\ 1\ \ 1\ ]^t$. This results in the following input-state-output history:

k	0	1	2	3	4	5	6
u(k)	1.0000	-.2360	0.3820	-.6180	1.0000	-1.618	2.6180
x(k)	1.0000	2.0000	-.2360	0.3820	-.6180	-1.0000	-1.6180
	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
y(k)	2.0000	2.0000	-.2360	0.3820	-.6180	1.0000	-1.6180

Observe that, because of the non-zero initial state, the observable excited dimension equals 2, despite the fact that only one state component is controlled by the input. Also observe that with $i = 3$ and $j = 5$, theorem 3 can not be employed now to analyse the possibility of rank cancellation. The reason is the non-zero initial state, which urges to employ lemma 1 in order to convert the system to an equivalent system, with zero initial state but with 2 inputs. However, in this case $2i = 6 > 5$, which violates one of the requirements of theorem 3. However, rank cancellation occurs in this example. Because, with:

$$U_h^{\perp} = \begin{pmatrix} 0 & 0 \\ 1/\sqrt{3} & 0 \\ -1/\sqrt{3} & 0 \\ -1/\sqrt{3} & 0.8506 \\ 0 & 0.5257 \end{pmatrix}$$

it can be verified that:

$$XU_h^{\perp} = \begin{pmatrix} 1.070 & 0 \\ 0 & 0 \end{pmatrix}$$

which is of rank 1. Hence, we have lost one dimension. The reason is of course that we have *tempted fate*, by choosing to small a j compared to i!

Example 4:

This examples aims at illustrating the several possible situations that may influence the rank of the concatenation of input and output block Hankel matrix. It provides a summary of the rank properties as previously discussed. Hereto consider the following system with 2 inputs, 1 output and 3 states:

$$A = \begin{pmatrix} \sqrt{3}/2 & -1/2 & 0 \\ 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & -0.92 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & -3 \end{pmatrix}$$

$$C = (\ 1\ \ 1\ \ 1\ ) \qquad D = (\ 0\ \ 0\ )$$

The initial state equals $x_0 = [\ 0\ \ 0\ \ 0\ ]^t$. The block Hankel parameters are $i = 6$ and $j = 19$. Hence U_h is a 12×19 while Y_h is a 6×19 block Hankel matrix. The system is identified

adaptively by a gliding window approach. The window is 24 time steps long: e.g. the first window extends from time 1 to 24, the second from 2 to 25, etc .... In each window the system is considered to be time invariant. The first identification result is obtained at time step 24. If we refer to a certain time step, this corresponds to the time step at which the identification is obtained, For instance, the identification at time 33, uses data from $k = 10$ to $k = 33$. The history of the system is the following:

Time step 24 - 40 : The first input is zero. The second one is a random signal, Gaussian distributed with mean zero and variance 1. The rank of the input block Hankel matrix is 6. The observable excited dimension is 1.

Time step 41 - 80 : A defect sensor changes the matrix C into $C = [\ 1\ 1\ 0\ ]$ at time 41. This makes the third state unobservable. Between timestep 41 - 64, the rank of the input-output block Hankel matrix reaches a maximum of 11. The reason is that the gliding window identification tries to model the timevariance of the system, by a linear time-invariant system but of higher complexity. Remarkable enough, for this example, it succeeds successfully: The singular values indicate that the behavior could be modelled by a linear time-invariant system of order 6. The transient phenomenon disappears from time step 64 on, because from then on, the data in the gliding window were generated exclusively by the non-observable time invariant system. Observe that the emerging and disappearing of the singular values proceeds rapidly.

Time step 81 - 120 : The first input is activated into a step-signal of amplitude 1 at time 81. This excites the states 1 and 2. Because the corresponding eigenvalues are on the unit circle, the step response is a linear oscillation. Observe that between time steps 81 and 86 the rank of the block input Hankel matrix increases up to its maximum, which is 12. The reason is that the step response gradually moves in into the input block Hankel as the window proceeds. This maximal rank persists until time step 99. At time 100, the rank starts to decrease again because of the constant first input. At time 105, the rank of the input block Hankel equals 7. Meanwhile, it is easily derived that the observable excited dimension equals 2 (there is no rank cancellation).

Time step 121 - 160 : Because a new sensor has been installed and activated, the system has again become observable at time step 121. However, the new sensor is not of high quality -despite its price!) : it shows drift which happens to be a first order autonomous unstable behavior with a pole at 1.01 (You have this with these cheap sensors!) The identified rank between 121 and 143 reflects the attempt of the identification algorithm to model the time variance (transition from unobservability to observability) by a higher complexity. However, this transient behavior disappears from time step 144 on. Between time 124 and 160, the rank of the concatenated input-output block Hankel matrix equals 11, which can be explained as follows: Rank 1 from the first input (a constant), rank 6 from the second input (random signal),observable excited dimension 2 of the oscillation, observable excited dimension 1 from the third state, which became observable again (the new sensor) and observable excited dimension 1 of the sensor's drift.

Time step 161 - 200 : In an attempt to stop the oscillation, an operator closes the open loop system with a static state variable feedback controller of the form $u_k = -Fx_k$, applied to the first 2 states and the first input, where $F = (\ -0.25\ \ 0.5\ )$. The controller succeeds in damping the oscillation. Again, between 161 and 183, there is a transient

rank increase and decrease due to time variance. From 184 on, the rank stabilizes again. However, rank cancellation for 2 dimensions of the first type now occurs, since the input is no longer independent. Observe that the drift is now very well recognizable in the output. The input-output rank now equals 10 = 6 (second input) + 2 (first input) + 4 (observable excited dimension) - 2 (rank cancellation from the state feedback).

8.2.6 The mathematical characterization of causal dependency

In the identifcation of large complex systems, it is not always clear *a priori* which of the variables are to be considered as inputs and which of them are outputs. One of the consequences of the rank property is, that *causal dependency* can be characterized mathematically via the singular values of the input-output block Hankel matrix. The idea behind this characterization is the fact that a variable is to be considered as an input variable if it is essentially independent of the other variables that have been observed. An input variable is a *free* variable. On the contrary, one of the essential features of an output is it dependency on the other variables. Since we are dealing with linear models, it follows that the investigation of dependency and independency reduces to analysing the rank of a matrix. It comes as no surprise if again the singular values are most appropriate to do the job. In those applications where the system is a complete black box, in such a way that it is not even known what are inputs and what are outputs, such a mathematical characterization is a most powerful tool. Note that it might even be unknown whether there are any inputs after all! Of course, the identification algorithm should be such that it recognizes if the system is *autonomous* or not. In a lot of applications however, it is known a priori which variables are inputs and outputs, in other words, the direction of the causal dependency relations is fixed. In these cases, the mathematical description of causal dependency may be used to analyse the effectiveness of the control strategy or of the design of the system. The discussion on rank cancellation has shown that one should take care with this a priori choice of causal dependency: Typically in a control environment, some of the so called inputs may become dependent upon the outputs. Despite the fact, that we keep calling them 'input variables', *strictu sensu*, they are no longer independent and hence, mathematically, are not longer to be qualified as inputs.

We shall now analyse in detail the detection of inputs in the mathematical sense: i.e. causal dependency will not be imposed or assumed a priori, but via an *algorithm.*

Theorem 4 *Given a sequence of $p \times 1$ data vectors w_k. Assume there exist indices i and j such that the block Hankel matrix W with i block rows and j block columns, with $j \geq pi$, constructed from the vectors w_k satisfies:*

$$r_W = rank(W) < p(i-1)$$

and that in addition, W satisfies the partial realization criterion. Then the data w_k , $k = 1, \ldots, i+j-1$ can be modelled by a dynamic linear time-invariant system.

Proof : The complete proof is not given here. It reduces to a combination of the rank property in theorem 2 and the results on realization theory of chapter 7. Moreover, a proof can also be found in [48] □

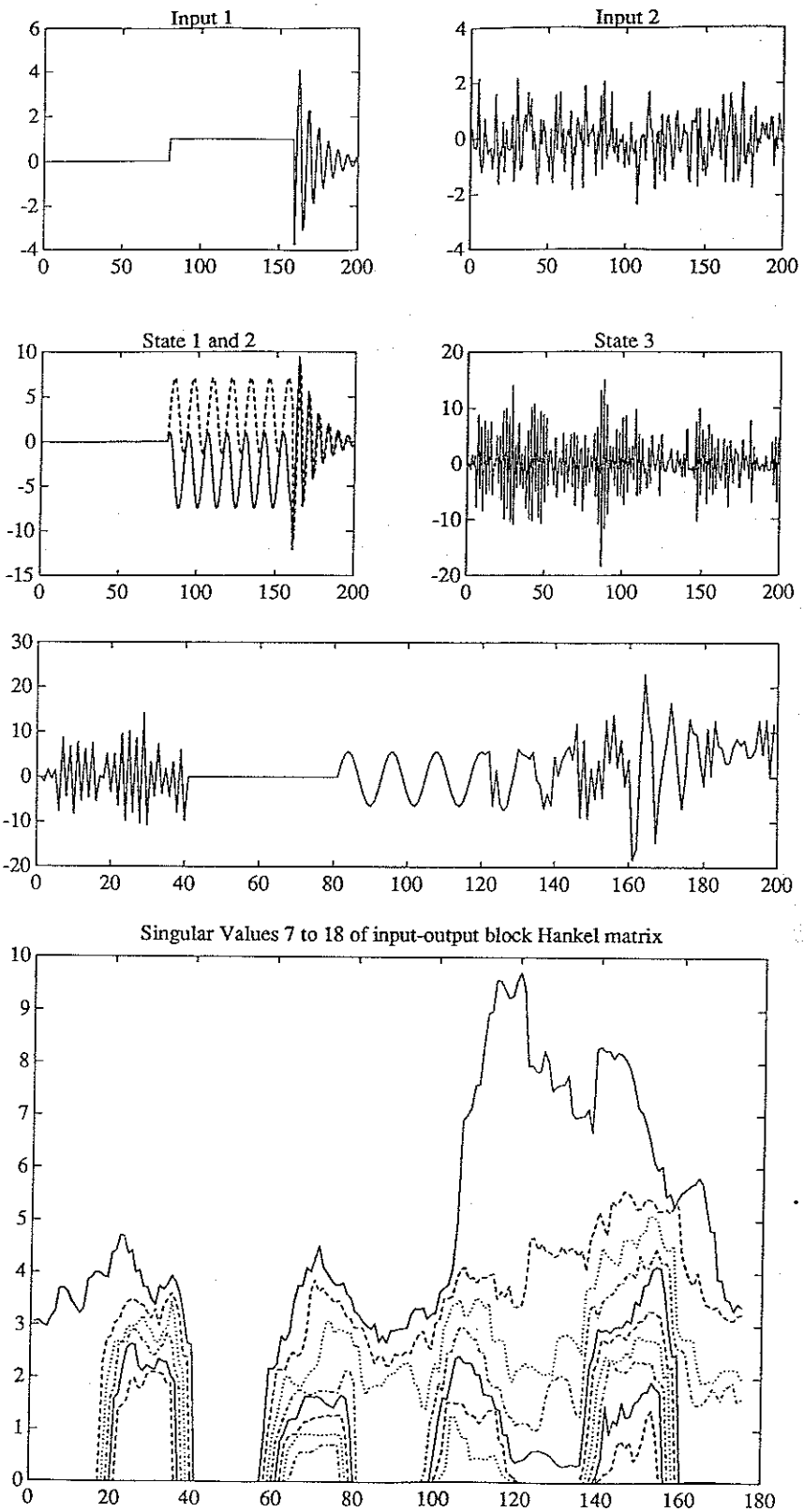


Figure 8.1: The inputs, the states, the output and singular values 7 to 18 of the input-output block Hankel matrix as a function of time for the timevarying system of example 4.

The rank condition $r_W < p(i-1)$ instead of pi is a consequence of the so called partial realizability criterion, which we have also encountered in chapter 7. The comparision with the results of chapter 7 provides moreover a remarkable interpretation of the contents of the theorem: Whenever the rank criterion is satisfied, the data can be modelled by a *linear time invariant autonomous linear system!* Hence, even if there are real inputs, that caused the behavior, mathematically, *it is even not needed to know them*, since the necessary rank condition allows to model the data as an autonomous linear system. This confirms the point of view, elaborated in [48], that 'causal dependency' is a matter of mathematics and not of subjective a priorisms!

Two extreme situations can be considered. Recall the realization algorithms based upon block Hankel matrices of chapter 7. These are autonomous linear systems (without physical input). Yet, they fit perfectly well in our framework as the rank property theorem will deliver the correct state space order. The other extreme occurs when the block Hankel matrix W is of full row rank. In this case, the p variables contained in the vector sequence w cannot be explained by something as *simple* as a linear system. Since any matrix is *generically* of full rank (already without Hankel structure), this fact confirms the observation that the occurrence of *linear time invariant dynamic systems*, and certainly of simple systems of low complexity (order smaller than i) is really rare in Nature. Moreover, this affirms the statement that *the heart of any identification approach is approximation!* If one wants to obtain a linear state space model, the full rank block Hankel matrix should be replaced with a rank deficient one. It confirms that the *inspiration* approach is in most cases to be preferred above the *deductive* approach.

Example : As an example, consider the following conversion table of temperature, expressed in degrees Fahrenheit to degrees Celsius.

^oC	^oF	^oC	^oF	^oC	^oF
5	41	25	77	45	113
10	50	30	86	50	122
15	59	35	95	55	131
20	68	40	104	60	140

Consider the vectors of pairs of related temperature:

$$w_k = \begin{pmatrix} T(^oF) \\ T(^oC) \end{pmatrix}$$

and construct a block Hankel matrix with as block dimensions $i = 3$ and $j = 9$. There are 4 non-singular values. Hence, the relation between these two variables can be modelled as a 4-th order autonomous linear system. However, observe that a 3×9 Hankel matrix constructed with only one of the variables, is of rank 3. This linear independence suggest that, one of the 2 temperatures can be considered as an input. The other one then becomes a dependent variable and can be considered as an output. Of course, all this follows also immediately from the well known conversion formula $T(^oF) = 9/5T(^oC) + 32$.

This example demonstrates that any of the 2 variables involved can be considered as an input. As a matter of fact, there is no *physical causal dependency* whatsoever involved.

This example in addition provides an interesting point of view on a linear relation between 2 variables when there is an *off-set*. If only observations on the 2 variables are given, the system that generated the data, should be considered as dynamical, having in its *memory* precisely the value of the off-set.

Despite the fact that mathematically, it is not needed to know the inputs in order to model data in a descriptive way (via an autonomous linear system), this approach is completely unsatisfactory when considered from the point of view of control. Here one tries to influence the behavior of a system by manipulating its free variables, which are called the inputs. Hence, it should be possible to detect which of the p variables of the vector sequence w are candidate inputs and which of them are to be considered as outputs. Two procedures might be developed:

1. Via an extensive number of rank tests on block Hankel matrices, employing the results on the rank of the input-output block Hankel matrix of theorem 2 and the fact that a block Hankel matrix containing only inputs, must be of full rank, it is possible to determine inputs and outputs. It cannot be too difficult to derive a systematic procedure.

2. In a second approach, the question of the determination of the causality relation, is postponed until the explicit computation of the system matrices. By a generalization of the conventional linear system equations, one can detect causality and anti-causality via the non-properness of the transfer function. This generalization goes via the introduction of a possibly singular matrix E and rewriting the state equation as:

   $$Ex_{k+1} = Ax_k + Bu_k$$

   A quick glance at the identification methodology to identify the model matrices, once an appropriate state sequence has been determined (see section 9.5.3), allows to conjecture that without much complication, one may also be able to include the identification of a singular matrix E. This conjecture however is still subject to further investigation.

Some other consequences of theorem 4, inspired by recent results obtained in [48], are summarized in the following corollaries:

Corollary 3 Model of a Linear System
Given a sequence of $p \times 1$ data vectors w_k. Assume that there exist indices i and j such that the block Hankel matrix W with i block rows and j block columns, with $j \geq pi$, constructed from the vectors w_k, satisfies:

$$r_W = rank(W) < p(i-1)$$

and that in addition W satisfies the partial realization criterion. Then, there exists a $(pi - r_W) \times pi$ matrix P of rank $(pi - r_W)$ such that:

$$PW = 0$$

Proof: Trivial. □

The matrix P might be considered as the model of the data that are contained in the matrix W. P is representative for a certain allowed behavior: all rank deficient block Hankel matrices,

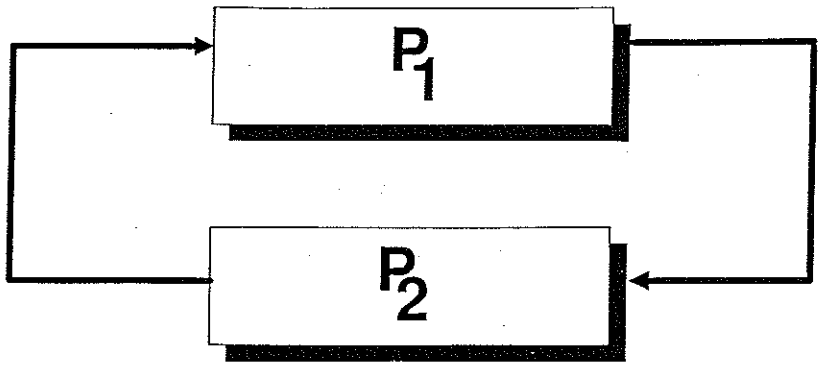


Figure 8.2: Output Feedback Configuration.

the column space of which is not orthogonal with respect to P's row space, are not compatible with the model. Observe that our analysis of the structure of block Hankel matrices in chapter 7, provides considerable insight into all possible data sequences that are compatible with a certain model P. It is possible to parametrize all rank deficient block Hankel matrices of which the column space is orthogonal to the rows of P. This could be done via the following algorithmic procedure:

1. Compute an orthonormal basis for the kernel of the matrix P.
2. Detect in this subspace the shift invariant subspace of maximal dimension with the structure exploiting factor analysis technique of section 7.4.
3. Compute the set of corresponding eigenvalues.
4. For any subset of these eigenvalues, one can construct rank deficient block Hankel matrices from a simple application of theorem 7 of section 7.3.1.
5. The data in any of these rank deficient block Hankel matrices are data compatible with the model P.

Observe that the matrix P is the mathematical translation of the notion of *falsification* as one of the basic underlying principles of mathematical models [48].
As a simple application of the preceding properties, consider the analysis of feedback behavior:

Corollary 4 Output Feedback Compatible Behaviors.
Assume that P_1 and P_2 represent models of dynamic systems, that are cascaded in an output feedback configuration. Then, any model P for the behavior that is compatible with this closed loop system, consists of any matrix P of which the rows are a basis of the intersection of the row spaces of P_1 and P_2.

$$span_{row}(P) = span_{row}(P_1) \bigcap span_{row}(P_2)$$

8.2.7 Identification of systems with delays.

The transfer function of a lot of physical systems may show a certain delay. Such a delay may have various physical causes. For instance, in a glass tube production process, the

delay is caused because the thickness measurements can not be performed accurately before the temperature of the tube has lowered beneath a certain threshold. Other causes include a priori filtering (e.g. anti-aliasing) which may introduce certain delays. In the transfer function of continuous time systems, a delay of T time units is modelled by a factor of the form e^{-Ts}, This introduces considerable mathematical complication because of the fact that a continuous time system with delay is in fact infinite dimensional. Fortunately, for a discrete time system, a delay can be very simply modelled by an appropriate state space model of an order which is determined by the ratio of the length of the delay with respect to the sampling time. In the sequel, it will be assumed that this is an integer number. For a more detailed analysis however, the reader is refered to [6].

Theorem 5 State space models for delays.
Consider a 'square' system with m outputs and m inputs with input-output equation given by: $y_k = u_{k-p}$. *Then, a minimal state space model is given by:*

$$\begin{pmatrix} u_{k-p+1} \\ u_{k-p+2} \\ \dots \\ u_{k-1} \\ u_k \end{pmatrix} = \begin{pmatrix} 0 & I_m & 0 & \dots & 0 & 0 \\ 0 & 0 & I_m & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & I_m \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} u_{k-p} \\ u_{k-p+1} \\ \dots \\ u_{k-2} \\ u_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ I_m \end{pmatrix} u_k$$

$$y_k = (\, I_m \, 0 \, \dots \, 0 \, 0) \begin{pmatrix} u_{k-p} \\ u_{k-p+1} \\ \dots \\ u_{k-2} \\ u_{k-1} \end{pmatrix}$$

Proof: Trivial. Observe that the system is minimal because the observability matrix Γ_n and Δ_n are given by:

$$\Gamma_n = \begin{pmatrix} I_m & 0 & \dots & 0 & 0 \\ 0 & I_m & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & I_m \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \qquad \Delta_n = \begin{pmatrix} 0 & 0 & \dots & I_m & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & I_m & \dots & 0 & 0 & \dots & 0 \\ I_m & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}$$

Both are of rank pm, hence the system is observable and controllable. Moreover, the conditioning of the controllability and observability matrix is perfect! □

The theorem allows to analyse in a straightforward way the effect of p delays on the rank of the corresponding input-output block Hankel matrix. Observe that the state vector sequence of the delay state space model generates a block Hankel matrix consisting of inputs:

$$X = \begin{pmatrix} u_{k-p} & u_{k-p+1} & \dots & u_{k-p+j-2} & u_{k-p+j-1} \\ u_{k-p+1} & u_{k-p+2} & \dots & u_{k-p+j-3} & u_{k-p+j} \\ \dots & \dots & \dots & \dots & \dots \\ u_{k-2} & u_{k-1} & \dots & u_{k+j-4} & u_{k+j-3} \\ u_{k-1} & u_k & \dots & u_{k+j-3} & u_{k+j-2} \end{pmatrix}$$

A closer look at this block Hankel matrix is that it provides an extension of the input block Hankel matrix over p time steps in the past. A simple application of the rank property reveals that the delay will be identifiable from the singular values if there is no rank cancellation between the row space of the input block Hankel matrix U_h, containing inputs $u_k, u_{k+1}, \ldots, u_{k+j-1}, \ldots, etc \ldots$ and the input block Hankel matrix formed with the inputs $u_{k-p}, \ldots, u_{k-1}, \ldots, u_{k+j-2}$. However, because of the block Hankel structure, this would only be possible if the input signal were structured in the sense that it be the output of an autonomous linear system. The other extreme is complete absence of such a structure, represented by an input consisting of white noise. Indeed, in this case, for large overdetermination j/i the block Hankel matrix will tend to be orthonormal with a rowspace which is orthogonal to the row space of the state vector sequence. This is one of the most favorable geometrical configurations for an accurate identification. In this case of white noise, the input block Hankel matrix will have a rank equal to mi and the state will cause the rank of the input output block Hankel matrix to be equal to $mi + mp$, if the block dimensions of the block Hankels are chosen sufficiently large. Moreover, all singular values will have the same order of magnitude.

One may however not forget that we have only discussed here pure delay systems. In a practical environment, system dynamics caused by the dynamical system under study are superposed on the pure delay singular values. The recognition of the delays, as zero eigenvalues of the system matrix (in discrete time), is however a difficult (if not impossible) task. The reason is the utmost sensitivity of eigenvalues with small modulus. This indicates that, if the delay can not be estimated from the singular values only, and then be 'eliminated' by an appropriate relative alignment of the corresponding input- and output sequences, it is in practice difficult to remove it.

Having now completed the development and discussion of the some general conceptual ideas, we shall now discuss several identification algorithms from a more pragmatic (deductive) point of view. They are based upon the fundamental input-output equation of theorem 1 and upon the rank property of theorem 2. For each strategy, we shall emphasize the essential features of method such as the geometrical observation on which it is based, the conceptual interpretation, (deductive - inspirationist), the imposed causal dependency assumptions, the algorithm etc.... The behavior of the algorithm is also investigated with respect to robustness and noise sensitivity.

8.3 Linear Least Squares Identification: Heuristics.

Starting from some basic geometrical observations, we shall show how an algorithm can be constructed to identify the system matrices A, B, C and D for a set of given data satisfying the following criteria:

- The obtained modelling strategy is *deductive*: It is assumed that the data were generated by a real existing finite-dimensional linear dynamical system. Except for this information (which is however already quite a lot!), all the rest is unknown (in casu the order and the system matrices).
- *Causal dependency* is imposed a priori: *Ab initio*, it is known (or assumed to be known), which variables are inputs and which are outputs.

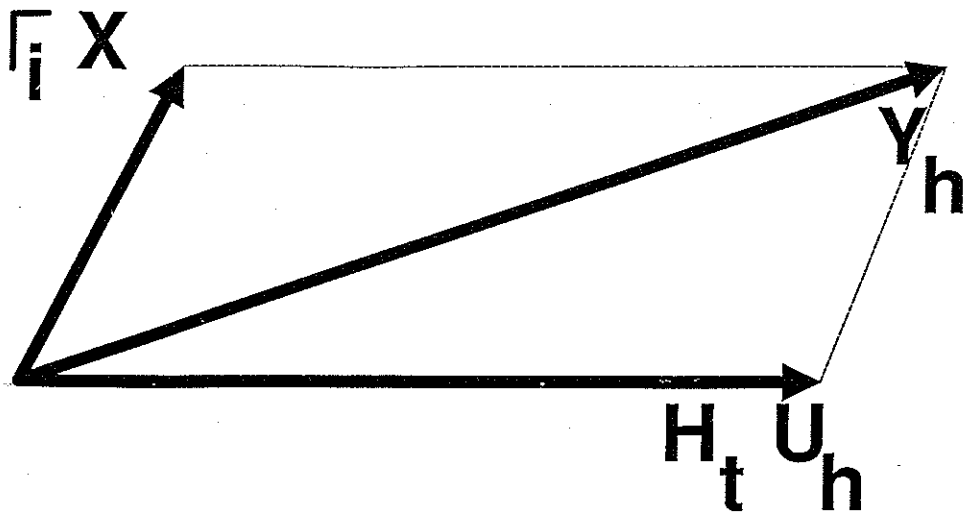


Figure 8.3: Geometrical representation of $Y_h = \Gamma_i X + H_t U_h$ in the row space

- Exact modelling is complicated by the fact that only noise corrupted *outputs* of the system are measurable. The noise is assumed to be additive. The *inputs* are assumed to be free.
- The qualification as *least squares* is only justified by analogy: Indeed, the main observation will be a *projection argument.* A criterion optimizing least squares strategy is the subject of section 8.5.

8.3.1 The geometrical observation

Two important geometrical observations will lead to an identification algorithm. First, everything is derived under the assumption that the data are noiseless. Then we analyse the effect of the additive perturbations, exploiting the *orthogonality* and the *lever* theorem of chapter 5. It is always assumed that the block dimensions i and j of the fundamental input-output matrix equation are sufficiently high as to ensure a meaningful interpretation of the proposed operations. The precise conditions will be derived furtheron.

The input-output equation in the row space

The geometrical interpretation of the matrix input-output equation when considered in the row space is most easily understood from the corresponding vector scheme in figure 8.3. Observe that the row space of Y_h is formed by linear combinations of row vectors of the the state vector sequence X and by linear combinations of vectors of the row space of U_h. *Assuming that there is no rank cancellation,* it is easily derived from the rank theorem 2, that the subspace, generated by the orthogonal projection of the rows of $\Gamma_i X$ onto the orthogonal complement of the row space of U_h, will be of dimension n_{oe} which is the observable excited dimension. Hence, this projection allows to state two important observations:

1. In the absence of rank cancellation, the observable excited dimension is the dimension of the projection of the row space of Y_h onto the orthogonal complement of the row space of U_h.

2. In the absence of rank cancellation, the observable excited dimension is the number of non-zero canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$.

3. In case there is rank cancellation, part of the row space of the state sequence X will be a subspace of the row space of the input block Hankel matrix U_h.

The following conditions on the block dimensions i and j are necessary in order to be able to do the interpretation in the row spaces:

1. $i_o < i$. Here i_o is the so-called observability index of the system. It is the minimum number of block rows that is needed so that $rank(\Gamma_i) = n_o$, where n_o is the dimension of the observable part of the state space. Note that $n_{oe} \leq n_o$.

2. $rank(U_h) \neq j$. If $rank(U_h) = j$, there would exist no orthogonal complement of the row space of U_h. Hence, we have as well $j > rank(U_h)$. Observe that, if there is no structure in the input (all inputs are independent variables), this implies that $rank(U_h) = mi < j$. Hence, the block Hankel matrix is rectangular with more columns than rows. In the sequel, this condition will always be true. Especially when the data are noisy, we shall find that $mi \ll j$ is absolutely necessary.

3. $j \geq rank(U_h) + n_{oe}$. The dimension of the orthogonal row complement of U_h equals $j - rank(U_h)$. If the dimension of the projection of the state sequence onto this orthogonal complement is to be n_{oe}, there should be enough 'space'. Hence $j \geq rank(U_h) + n_{oe}$.

Summarizing, a safe choice for arbitrary inputs is:

$$n_o \leq i \leq mi \leq (mi + n_{oe}) \leq j$$

Of course, since the observable dimension n_0 and the observable excited dimension n_{oe} are not known a priori, one should content oneself with sufficiently large overestimations. We shall assume from here on, that these inequalities are *at least* satisfied with equality.

The input-output equation in the column space

For the ease of discussion, it is assumed that $rank(U_h) = mi$ which implies that the inputs are not structured. Let $U_h^\perp$ be a $j \times (j - mi)$ orthonormal matrix such that:

$$U_h U_h^\perp = 0$$

Consider the output Hankel matrix Y_h and multiply it on the right with $U_h^\perp$. Then from the input-output equation, it follows immediately that:

$$Y_h U_h^\perp = \Gamma_i X U_h^\perp$$

Then it is easily seen that each column of the matrix on the left hand side, can be considered as the output of an autonomous linear system with as initial state the corresponding column of the matrix $XU_h^\perp$. This observation will be crucial in the derivation of an algorithm to retrieve the system matrices A and C. Observe that the row space of the matrix $Y_h U_h^\perp (U_h^\perp)^t$ represents the projection of the row space of Y_h onto the orthogonal complement of $span_{row}(U_h)$. This observation will lead to a *least squares* interpretation when dealing with

noisy outputs.

When there is rank cancellation, in the projection we shall lose that part of the system, of which the row space of the state sequence, belongs to the row space of the input block Hankel. Remember that in corollary 2, we have given a kind of perfect input sequence with respect to rank cancellation. It consisted of applying a Dirac impulse to all inputs at the same time. However, observe now that this input is also optimal with respect to the projection approach presented here. If the initial state equals zero, the row space of the state sequence will be orthogonal to the row space of the input block Hankel.

The projection interpreted in the frequency domain.

The Discrete Fourier Transform (DFT) X of an $1 \times k$ vector sequence x may be represented by a post-multiplication of x with a unitary $k \times k$ matrix W_k. When applied to the input-output equation of a SISO system, this results in:

$$Y_h W_j = \Gamma_i X W_j + H_t U_h W_j$$

where now the W_j is an $j \times j$ unitary DFT matrix. Though not generally true, the following remarks provide some useful *rules of thumb* for most practical systems and input signals, that are well suited for identification:

- A favorable identification requirement is the fact that the frequency range of the input sequence, is larger than the pass band of the system. The reason is of course the *persistancy of excitation* condition, requiring that all modes of a system should be sufficiently well excited. Moreover, most real systems are relatively *low pass.*

- These frequency properties of systems have a direct effect on the input-output equation. First note that from the orthonormality of the DFT matrix W_j, the geometrical interpretation in the row space is completely preserved, in particular the canonical angles between $span_{row}(Y_h)$ and $span_{row}(U_h)$ are exactly the same as those between $span_{row}(Y_h W_j)$ and $span_{row}(U_h W_j)$. A row vector y of $Y_h W_j$ can be decomposed as a direct sum of a vector y_1 in $span_{col}(W_j U_h^{\perp})$ and a vector y_2 in $span_{row}(U_h W_j)$. Observe that the vector y_1 originates in the state sequence, hence will consist mainly of low frequency contributions, because of the low pass character of the system. The frequency content of the vector y_2 is determined by the input sequence frequency range and will consist of higher frequencies than those in y_1. If the vector y is almost orthogonal with respect to $span_{row}(U_h)$, then it will consists mainly of low frequencies. If it is close to $span_{row}(U_h)$, its frequency content will more ressemble that of the input. One can conclude that the frequency content of $span_{row}(U_h) \bigcap span_{row}(Y_h)$ will happen at high frequencies, relative to the system's pass band. This conclusion is confirmed in figure 8.4. Figure 8.4.a. shows the power spectrum of a vector of $span_{row}(Y_h)$ that makes an angle of 84° with $span_{row}(U_h)$, while figure 8.4.b. shows the power spectrum of a vector of $span_{row}(Y_h)$ within $span_{row}(U_h)$. The data used are those of the example in section 8.3.3.

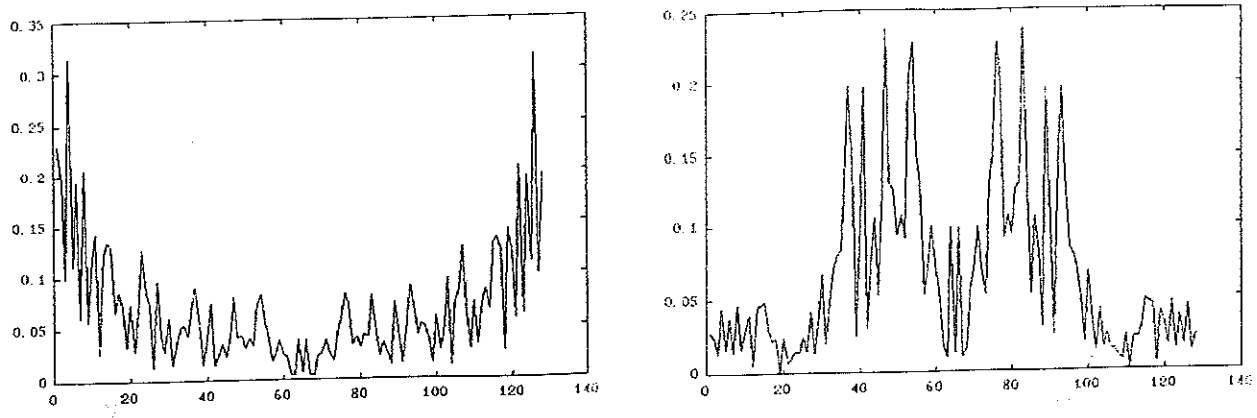


Figure 8.4: Power spectrum of a vector of $span_{row}(Y_h)$ almost orthogonal to $span_{row}(U_h)$ and a vector within $span_{row}(U_h)$.

8.3.2 The mathematical observation

It is easily verified that the problem of identifying a state space model is linear in the matrices B, C and D but non-linear in the system matrix A. It will now be shown how the matrices A and C can be obtained by exploiting the shift structure of the projected output block Hankel matrix in a completely similar manner as in Kung's algorithm (chapter 7). Once A and C are determined, the matrices B and D follow from the solution of a set of linear equations.

Computation of the matrices A and C.

Reconsider again the expression for the projected output block Hankel matrix:

$$Y_h U_h^{\perp} = \Gamma_i X U_h^{\perp}$$

Two important observations allow to compute the matrices A and C:

Theorem 6 Identification of A and C.
Assuming that:

1. *Y_h is the $li \times j$ output block Hankel matrix and U_h the $mi \times j$ input block Hankel matrix of a linear system with observable dimension n_o and observable excited dimension n_{oe}.*
2. *$i_o \leq (i+1) \leq (mi + n_{oe}) \leq j$ where i_o is the observability index of the system.*
3. *$U_h^{\perp}$ is an $j \times (j - mi)$ orthonormal matrix such that $U_h U_h^{\perp} = 0$.*
4. *The singular value decomposition of $Y_h U_h^{\perp}$ is given by:*

$$Y_h U_h^{\perp} = (P_1 \; P_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} Q^t$$

with $rank(S_1) = r$ and P_1 has r columns.

5. *There is rank cancellation for r_{canc} dimensions*
6. *$\underline{P_1}$ ($\overline{P_1}$) is the $l(i-1) \times r$ matrix obtained from P_1 by omission of the last (first) block row. p_1 is the first block row of P_1.*

then:

1. *$rank(S_1) = r = n_{oe} - r_{canc}$.*
2. *Models for the system matrices corresponding to the part of the system that is not affected by rank cancellation, are given by $C = p_1$ and A is the total linear least squares solution of $\underline{P_1}A = \overline{P_1}$.*

Proof: The rank property follows immediately from the geometrical interpretation in the row and column spaces. The identification of A and C follows from the fact, proved in chapter 7, that the shift structure is an intrinsic property of a subspace and does not depend upon the choice of basis in that subspace. It is easily verified that the left singular vectors contained in the matrix P_1 generate a basis for the column space of Γ_i. Observe that we have required that $i_o \leq (i+1)$ instead of $i_o \leq i$. The reason is to be found in the partial realization criterion, which was also explored in chapter 7. □

- If there is no rank cancellation, the theorem provides a nice algorithm to compute directly the matrices A and C from the shift structure of the projection of the output block Hankel matrix. Despite the elegancy for the computation of A and C, the computation of B and D is more elaborate.
- Observe that rank cancellation occurs if:

$$rank(XU_h^{\perp}) < n_{oe}$$

- In the theorem, we use the total linear least squares solution for A. Obviously this is due to the fact that both $\underline{P_1}$ and $\overline{P_1}$ are subject to approximatively the same amount of noise.

Computation of the matrices B and D.

Observe that, once the system matrices A and C have been found, the identification problem is linear in the matrices B and D. They can be solved from a set of linear equations as is now demonstrated. The result will be derived for the case that $rank(U_h) = mi$, but could easily be generalized if U_h were rank deficient. Premultiply the input-output block Hankel matrix with a $(li-n) \times li$ matrix of rank $li-n$, which satisfies $Q\Gamma_i = 0$. Observe that in theorem 6, such an matrix has already been found in the SVD of the matrix $Y_h U_h^{\perp}$. This results in:

$$QY_h = QH_tU_h$$

If this equation is post-multiplied with the pseudo-inverse U_h^+ of the matrix U_h, one finds:

$$QY_hU_h^+ = QH_t$$

Observe that the required pseudo-inverse U_h^+ is already available from theorem 6. Now call K the $(li-n) \times mi$ matrix $QY_hU_h^+$. Then, introducing an obvious partitioning of K and Q, one can write:

$$(K_1\ K_2\ \ldots\ K_i) = (Q_1\ Q_2\ \ldots\ Q_i) \begin{pmatrix} D & 0 & 0 & \ldots & 0 \\ CB & D & 0 & \ldots & 0 \\ CAB & CB & D & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ CA^{i-2}B & CA^{i-3}B & \ldots & \ldots & D \end{pmatrix}$$

This equation can be rewritten into a set of equations of the from:

$$\begin{pmatrix} K_1 \\ K_2 \\ K_3 \\ \ldots \\ K_i \end{pmatrix} = \begin{pmatrix} Q_1 & Q_2 & Q_3 & \ldots & \ldots & Q_i \\ Q_2 & Q_3 & Q_4 & \ldots & Q_i & \ldots \\ Q_3 & Q_4 & \ldots & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ Q_i & \ldots & \ldots & \ldots & \ldots & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & \Gamma_{i-1} \end{pmatrix} \begin{pmatrix} D \\ B \end{pmatrix}$$

The condition for this set of equations to have a unique solution, is:

$$(li-n)i \geq (l+n)$$

The following remarks apply:

- First, observe that it is not necessary to choose precisely $li-n$ linear independent vectors, which are the rows of Q, orthogonal to $span_{col}(\Gamma_i)$. Less, say q, may be sufficient as long as the overdetermination is satisfied as to guarantee a unique solution:

$$qi \geq (l+n)$$

- The state sequence can be computed from:

$$X = \Gamma_i^t(Y_h - H_tU_h)$$

  when Γ_i is orthonormal as in theorem 6.

- It is not necessary that Y_h and U_h are block Hankel matrices. Columns may be eliminated but each column must consist of i consecutive samples. This will also be proved furtheron in corollary 5.

- Two other methods to compute the matrices B and D are described in [42]. However, of the three methods developed in [42], the one presented here is the fastest and requires only an additional solution of a set of linear equations, which may be done e.g. with total least squares.

- Observe that in the computation it is important to use the matrix Γ_i as estimated from the SVD in theorem 6, as such, despite the fact that it may show no shift structure whatsoever. It is of no use to try to iterate with the derived models for A and C in order to form a matrix with a correct shiftstructure. The reason is that, when this approach would be applied, little errors obtained in the estimation of A and C may have a big

influence on the estimation of the remaining parameters.

Example:
Assume that a signal y_k is generated according to the law: $y_k = cos(\omega k)\alpha + sin(\omega k)\beta$, $k = 1, \ldots$ Now assume that an estimate $\omega + \delta\omega$ has been obtained and that now the parameters α and β are solved via a set of *linear* equations of the form:

$$(\; cos((\omega + \delta\omega)k) \;\; sin((\omega + \delta\omega)k) \;) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \simeq y_k$$

Observe that, as a first order approximation:

$$\begin{aligned} cos((\omega + \delta\omega)k) &\simeq cos(\omega k) + \frac{dcos(\omega k)}{d\omega}\delta\omega = cos(\omega k) - ksin(\omega k)\delta\omega \\ sin((\omega + \delta\omega)k) &\simeq sin(\omega k) + \frac{dsin(\omega k)}{d\omega}\delta\omega = cos(\omega k) + kcos(\omega k)\delta\omega \end{aligned}$$

Observe that the noise is not independent of the 'exact' variables, though on the other hand, the qualitative structure of the noise matrix is approximately known. Yet note that the noise energy increases with time, indicating that the more equations we use, the larger will be the deviation from the exact dynamics. Hence, the problem is difficult to solve so that it is not easy to find an appropriate estimate of α and β.

8.3.3 Deductive analysis of the influence of additive noise.

The qualification of the presented approach as 'least squares' is only by analogy: Indeed, the main observation is a projection argument as is the case in a linear least squares framework. However, the projection will be of full dimension li. Therefore, estimation of the order is crucial. When the order is estimated, it is finally shown that the lever theorem will allow to demonstrate the concistency of the needed shift structure.

Estimation of the observable excited dimension

First, we shall analyse the possibility of estimating correctly the observable excited dimension when the outputs are noisy. Hereto, we shall exploit the insights that we have developed in chapter 5 on the behavior of singular values and canonical angles of matrices when perturbed by additive noise. Some rules of thumb will be derived for an appropriate choice of the block dimensions i and j .

In order to illustrate our statements, we shall use the following $5 - th$ order system:

$$A = \begin{pmatrix} 0.9 & 0.3 & 0 & 0 & 0 \\ -0.3 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & -0.6 & 0.5 & 0 \\ 0 & 0 & -0.5 & -0.6 & 0 \\ 0 & 0 & 0 & 0 & 0.65 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & -1 \\ -1 & 4 & 1 \\ 1 & -1 & 2 \\ 3 & 0 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & -2 & 0 & 1 & 3 \\ -1 & 1 & 1 & 2 & -2 \end{pmatrix} \qquad D = \begin{pmatrix} 3 & 5 & 4 \\ 1 & 2 & 3 \end{pmatrix}$$

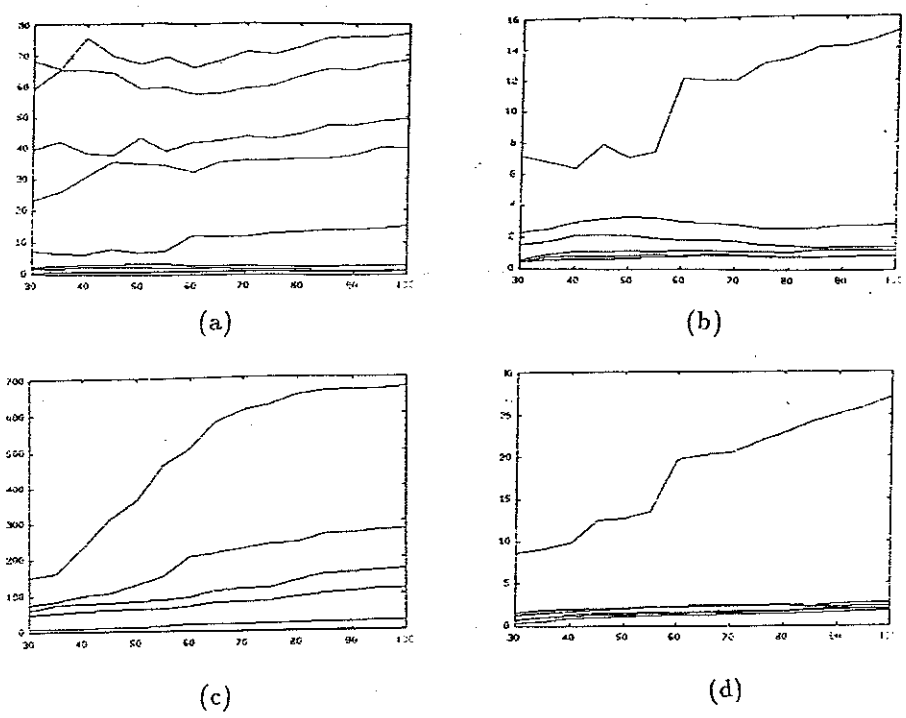


Figure 8.5: (a) The 10 canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$. (b) 6 smallest angles (c) The 10 singular values of $Y_h U_h^{\perp}$ (d) The 6 smallest singular values.

The applied input sequences are normally distributed noise with variance 1 and zero mean. The additive noise superposed on the output was 1 %.
In figure 8.5.a., we have depicted the canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$ as a function of j, for $i = 5$. An enlarged figure, showing only the 6 smallest canonical angles is found in 8.5.b. The qualitative conclusions of chapter 5 are obviously confirmed: The canonical angles that ought to be zero, and the canonical angles that are non-zero for exact data, both reach a threshold level as is predicted by theorem 12 of section 5.4.4. The 10 singular values of $Y_h U_h^{\perp}$ are depicted in figure 8.5.c. while only the smallest 6 ones can be found in figure 8.5.d. Observe that our analysis in chapter 5, predicted an increase as $\sqrt{j}$ of the singular values as a function of j. All singular values, both from the noise and the exact signal, obey this law.

Let's now analyse what happens when the ratio j/i is kept constant, and j is increased. The same example was used, but now with 5 % additive noise on the output. In figure 8.6.a, one finds the $2i$ canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$ as a function of i for constant $j/i = 3$ and in figure 8.6.b., the $2i - 4$ smallest angles. Again, the canonical angles all reach a kind of asymptotic level. Observe that it is now more difficult to estimate the correct order. The singular values of $Y_h U_h^{\perp}$ can be found in figure 8.6.c and 8.6.d. Note that, as a function of i with j/i constant, the first 5 singular values keep increasing, while the singular values caused only by the noise, barely grow. An explanation is the following: The energy in Y_h increases as i^2 for increasing i when j/i is kept constant. For exact data, the total energy in Y_h, along the $n = 5$ directions that are lying in the orthogonal complement of $span_{row}(U_h)$, will be distributed over these $n = 5$ directions. Hence, the 'exact' singular values will increase proportionally to i as i increases. When projected on the orthogonal complement of $span_{row}(U_h)$, the energy caused by the additive noise will be distributed over $li = 2i$ directions. Since the noise energy is proportional to i^2, each direction will receive an

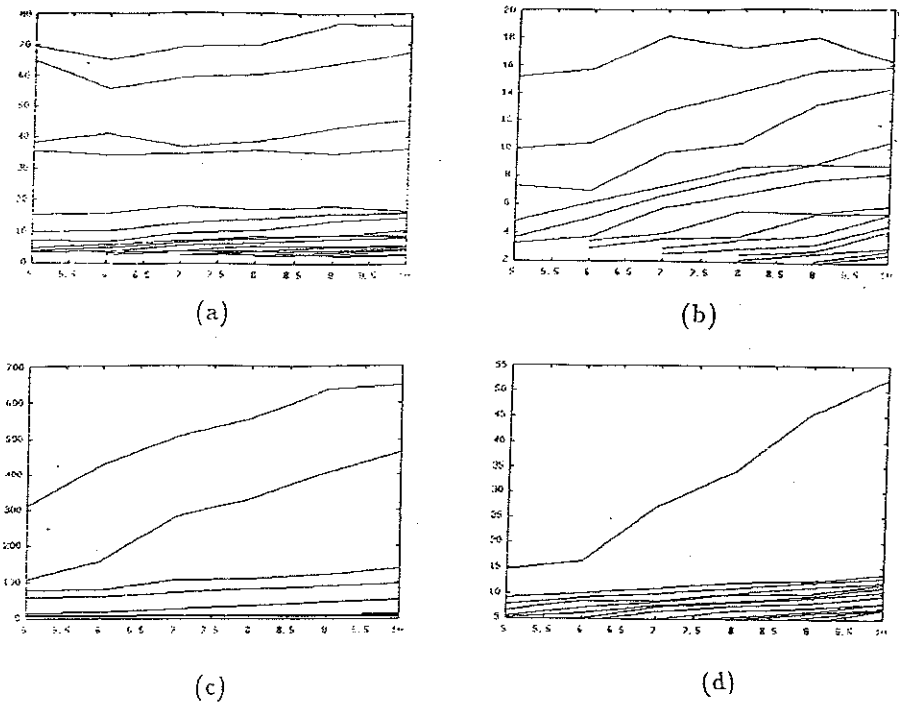


Figure 8.6: (a) $2i$ canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$. (b) $2i-4$ smallest angles. (c) $2i$ singular values of $Y_h U_h^{\perp}$. (d) $2i-4$ smallest singular values.

average energy of i^2/li, hence each singular value caused by the noise, will only grow as $\sqrt{i}$ as a function of i, which is slower that the linear growth of the exact singular values.

With respect to the choice of the block dimensions i and j, one can draw the following conclusions:

- The noise sensitivity of the order estimation, decreases for constant i, with increasing j. Hence j should be chosen large with respect to i.
- The singular values are much better to estimate the order than the canonical angles. The reason is due to a normalization when using canonical angles, which throws away all the information on the oriented energy of both the exact signals and the additive noise. However, it is precisely this oriented energy which forces the 'exact' singular values to grow much faster than those caused only by the noise. This phenomenon is directly connected to an increase of i while keeping j/i constant.
- Observe that when the total number of data N is constant, then i and j must be such that $N = i + j - 1$. Hence, there will be a kind of *compromis* between the choice of the optimal i and j, taking into account the phenomena just described.

Notwithstanding the fact that the least squares identification approach is in theory only suited for noise free inputs, it is interesting to investigate what would happen if also the inputs are noise corrupted. A remarkable constatation is that: *The order estimation is relatively more sensitive to noise on the outputs that noise on the inputs.* In order to illustrate this observation, reconsider the previous example. Figure 8.7. demonstrates the influence on the canonical angles between $span_{row}(U_h)$ and $span_{row}(Y_h)$ for $i = 5, j = 30$ as a function of the relative noise level and figure 8.8. does the same for the singular values. In figure 8.7.a., 8.7.b., 8.8.a. and 8.8.b., the inputsequences were disturbed, while in 8.7.c., 8.7.d, 8.8.c.and 8.8.d. the

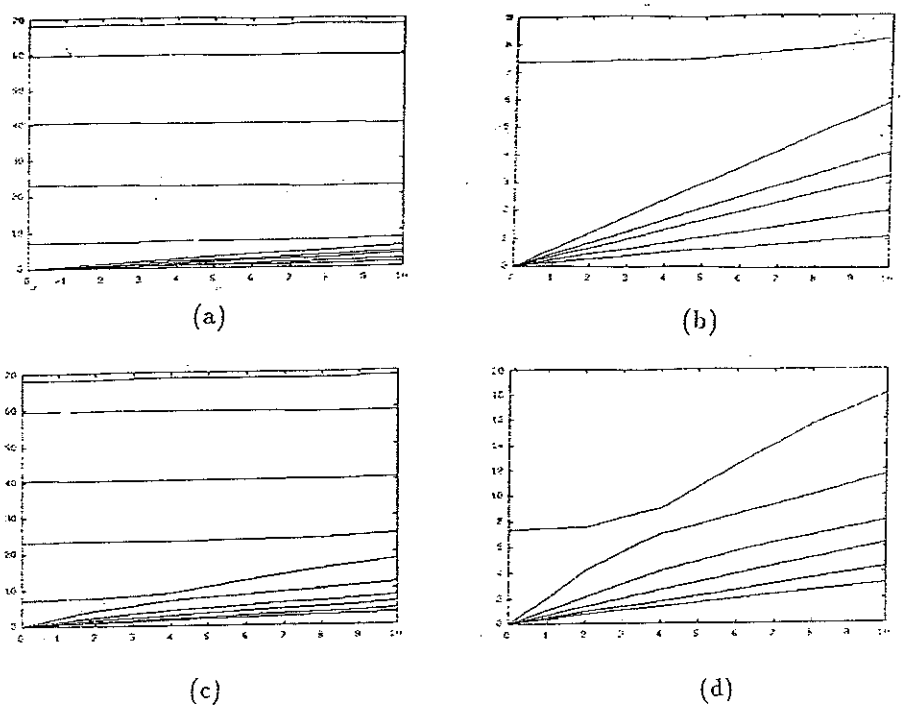


Figure 8.7: (a) 10 canonical angles as function of noise level on input. (b) 6 smallest angles. (c) 10 canonical angles as function of noise level on output (d) 6 smallest angles.

output sequences were noisy. This difference in noise sensitivity between inputs and outputs can be explained in terms of the frequency behavior of a system in relation to the fundamental input-output equation. Recall that the intersection between $span_{row}(U_h)$ and $span_{row}(Y_h)$ was mainly determined by the higher frequencies. Since the power spectrum of the output of a linear system is low-pass, relative to the flat power spectrum of the additive (white) noise, the noise will mainly affect the higher frequencies of the output. In the example, the frequency band of the input sequence was wide with respect to the pass band of the system. Hence, $span_{row}(U_h)$ will not be perturbed as much as $span_{row}(Y_h)$. The projection of Y_h onto the orthogonal complement of $span_{row}(U_h)$ however, is the complement of the intersection of $span_{row}(Y_h)$ with $span_{row}(U_h)$ within $span_{row}(Y_h)$. Hence, this projection will be more affected by noise on the outputs than noise on the inputs!.

Consistent estimation of the shift structure

Recall that the identification of the matrices A and C is possible from the shift structure of the column space of the matrix $Y_h U_h^{\perp}(U_h^{\perp})^t$. This matrix will however in general be of full rank when the outputs are noisy. However, having established in the preceeding section how to find a good estimate for the observable excited dimension, it is easily seen that a consistent estimation of the column space, and hence of the associated shift structure, will be possible by choosing j as large as possible. This follows directly from the lever theorem and its corollaries discussed in chapter 5. Observe moreover that the set of equations to be solved for A, is solved via *total linear least squares*, hence a second noise filtering effect will enhance the estimate of A, which will ameliorate if i increases, since the total linear least squares solution is consistent for increasing overdetermination, which in this approach depends on i.

As an example of the accuracy of the least squares algorithm, recall the previously defined

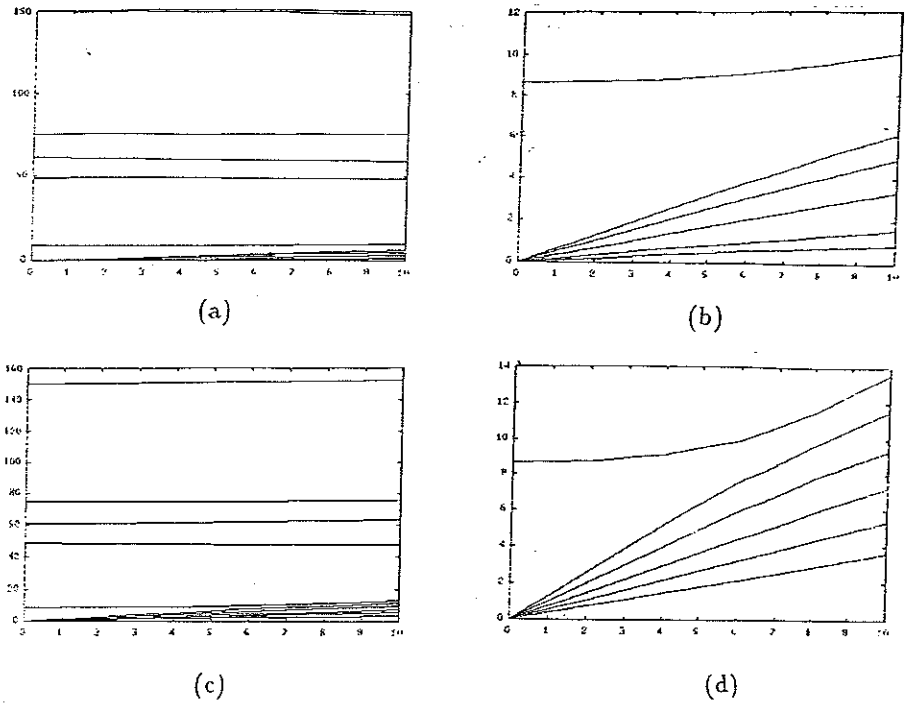


Figure 8.8: (a) 10 singular values as function of noise level on input.(b) 6 smallest singular values. (c) 10 singular values a function of noise level on output (d) 6 smallest singular values.

5-th order system with 1 % additive noise superposed on the outputs. From the lever theorem. it is expected that the column space of the matrix $Y_h U_h^{\perp}$, corresponding to the first 5 left singular vectors of $Y_h U_h^{\perp}$, will be a consistent estimate for the column space of the extendent observability matrix Γ_i. Observe that $span_{col}(\Gamma_i)$ is uniquely determined by the observable poles of the system as was proved in chapter 7. The consistency is demonstrated in figure 8.9., where the 5 canonical angles between the column space generated by the first 5 left singular vectors of $Y_h U_h^{\perp}$ and $span_{col}(\Gamma_i)$ are depicted for constant $i = 5$ and increasing j. Again, it may be observed that the accuracy of the identification is more sensitive in case of noisy outputs than in the case of noisy inputs. As an example, compare the canonical angles between $span_{col}(Y_h U_h^{\perp})$ and $span_{col}(\Gamma_i)$ in figure 8.10.a. as a function of the noise level on the inputs, and in figure 8.10.b. on the outputs.($i = 5, j = 30$).

8.4 Total Linear Least Squares: Heuristics.

Having discussed in detail in the previous section, the linear least squares solution, we shall now analyse an analog to the total linear least squares approach of sets of linear equations, applied to dynamic system identification. The main features of this approach are:

1. The modelling strategy is essentially *deductive*. It is assumed that the data originate from a linear time invariant system, but that they are corrupted by additive noise. Further it is assumed that both the noise on the input and output are white and of the same absolute variance. Mind that the *whiteness* condition is now essential in contrast with the case of solving unstructured linear equations. The reason is the (block) Hankel structure of the matrices. If however this structure would be removed, which is allowed by a simple application of corollary 5 to be proved in section 8.5.1, the noise assumptions may be considered less restrictive.

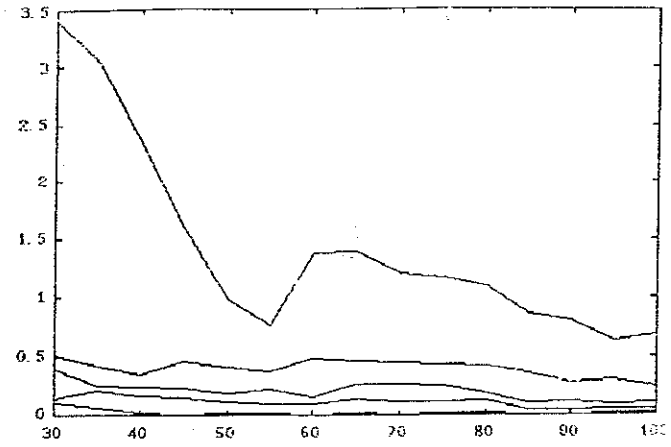


Figure 8.9: Canonical angles between $span_{col}(Y_h U_h^{\perp})$ and $span_{col}(\Gamma_i)$.

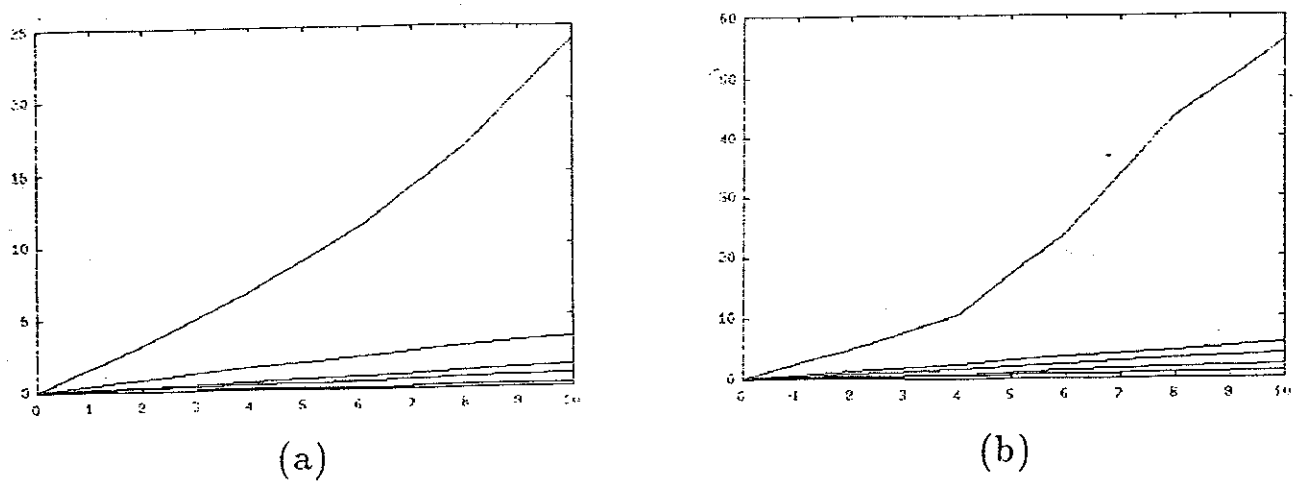


Figure 8.10: Canonical angles between $span_{row}(Y_h U_h^{\perp})$ and $span_{col}(\Gamma_i)$ as a function of the noise level (a) on the inputs. (b) on the outputs.

2. Causal dependency assumptions are imposed a priori.

3. Approximately the same deductive conclusions as in the least squares algorithm, apply for the behavior of the singular values and the canonical angles as a function of the block dimensions and the noise level.

4. The *total least squares* qualification is mainly inspired by *analogy* with the solution of sets of linear equations. Indeed, it will be shown how the model can in essence be derived from a basis of a certain subspace of the short space. A more rigorous, criterion minimizing total least squares approach will be the subject of section 8.5.2.

5. The presented results were derived by Lieven Vandenberghe and Piet Van Mieghem in their master's thesis [42].

In order to avoid an unnecessary complicated derivation, it is assumed that no rank cancellation occurs. If it occurs, then the order of the identified system will differ from the *real* system by an amount equal to the number of cancelled states.
Under these conditions, we have already shown that:

$$rank \begin{pmatrix} Y_h \\ U_h \end{pmatrix} = rank(U_h) + n_{oe}$$

Here n_{oe} is the observable excited order. Observe that this rank property allows to estimate the observable excited order of the system. Recall from the discussion on total linear least squares approach in chapter 5, that the matrices are first 'corrected' in order to find an approximation which is rank deficient. Vectors in the kernel of this rank deficient corrected matrix then serve as an approximation of the 'exact' linear relation. When a Frobenius-norm criterion is proposed, the solution is most elegantly delivered in terms of the singular value decomposition. It is now shown how a similar approach applies to dynamic systems. As is always the case in the derivation of deductive modelling strategies, we shall develop the algorithm in a noisefree context. Exploiting the knowledge on the total least square algorithm and the assumed properties on the noise, it will then be verified how the approach behaves when the data are noisy.

Theorem 7 Heuristic Total Linear Least Squares Identification.
Asssume that $rank(U_h) = r_u$, *there occurs no rank cancellation and that the column space of the state sequence coincides with the observable part of the state space* ($n_{oe} = n_o$), *and that the singular value decomposition of the concatenation of input and output block Hankel is given by:*

$$\begin{pmatrix} Y_h \\ U_h \end{pmatrix} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} (Q_1 \ Q_2)^t = (P_1 \ P_2) \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix}$$

where the matrix of left singular vectors is partitioned according to the dimensions of S_1 *and of* Y_h *and* U_h. *Then:*

1. $rank(S_1) = n_o + r_u$ *and* P_{11} *is a* $li \times (r_u + n_o)$, P_{12} *is* $li \times (li + mi - r_u - n_o)$, P_{21} *is* $mi \times (r_u + n_o)$, P_{22} *is* $mi \times (li + mi - r_u - n_o)$.

2. *There exists a non-singular $n_o \times n_o$ matrix T_1 and a non-singular $((l+m)i-(n_o+r_u)) \times ((l+m)i-(n_o+r_u))$ matrix T_2 such that:*

$$\begin{pmatrix} P_{11}P_{21}^{\perp} & P_{12}T_2 \\ 0 & P_{22}T_2 \end{pmatrix} = \begin{pmatrix} \Gamma_i T_1 & \Gamma_i^{\perp} \\ 0 & -H_t^t \Gamma_i^{\perp} \end{pmatrix}$$

where Γ_i is the extended observability matrix, the $(li-n_o)$ columns of $\Gamma_i^{\perp}$ are linear independent vectors from the orthogonal complement of $span_{col}(\Gamma_i)$, and H_t is the $li \times mi$ lower triangular block Toeplitz matrix with Markov parameters. The columns of $P_{21}^{\perp}$ are a basis for the orthogonal row complement of P_{21}.

Proof: The rank property has already been proved in theorem 2. First we prove that there exists a non-singular $n_o \times n_o$ matrix T_1 such that:

$$P_{11}P_{21}^{\perp} = \Gamma_i T_1$$

From the input-output matrix equation it follows that:

$$\begin{aligned}
(I \quad -H_t)\begin{pmatrix} Y_h \\ U_h \end{pmatrix} &= \Gamma_i X \\
(I \quad -H_t)P_1 S_1 Q_1^t &= \Gamma_i X \\
(I \quad -H_t)\begin{pmatrix} P_{11} \\ P_{21} \end{pmatrix} &= \Gamma_i X Q_1 S_1^{-1} \\
(I \quad -H_t)\begin{pmatrix} P_{11} \\ P_{21} \end{pmatrix} P_{21}^{\perp} &= \Gamma_i X Q_1 S_1^{-1} P_{21}^{\perp} \\
P_{11}P_{21}^{\perp} &= \Gamma_i (X Q_1 S_1^{-1} P_{21}^{\perp})
\end{aligned}$$

The columns of the matrix $Q_1 S_1^{-1} P_{21}^{\perp}$ are a basis for the n_o-dimensional part of the orthogonal complement of $span_{row}(U_h)$ in which lies the projection of Y_h. This can be seen from that fact that on the one hand, the columns of Q_1 are a basis for the sum of the row spaces of Y_h and U_h (this follows from the SVD) and on the other hand:

$$U_h Q_1 S_1^{-1} P_{21}^{\perp} = 0$$

If there is no rank cancellation, it follows that:

$$rank(X Q_1 S_1^{-1} P_{21}^{\perp}) = rank(Q_1 S_1^{-1} P_{21}^{\perp}) = n_o$$

But this implies that the column spaces of $P_{11}P_{21}^{\perp}$ and Γ_i coincide which proves the existence of a non-singular $n_o \times n_o$ transformation matrix T_1.
Premultiplication of the input-output equation with $(\Gamma_i^{\perp})^t$ yields:

$$\left[(\Gamma_i^{\perp})^t \quad -(\Gamma_i^{\perp})^t H_t\right] P_1 S_1 Q_1^t = 0$$

Hence the row space of the first matrix must be orthogonal to the column space of P_1, hence must be a linear combination of the columns of P_2, which proves the existence of a $((l+m)i-(n_o+r_u)) \times ((l+m)i-(n_o+r_u))$ non-singular transformation matrix T_2. □

While the formulation of this theorem looks quite complicated as well as its proof, it provides us with an utmost straightforward identification approach in the case that all data are noise corrupted.

- The observable excited order can be estimated from the singular values of the concatenation of input-output block Hankel matrices.

- The matrices A and C can readily be computed from the shift structure of the matrix $P_{11}P_{21}^{\perp}$ as in Kung's realization algorithm, explained in chapter 7.

- From the equality:

$$P_{22}T_2 = -H_t^t\Gamma_i^{\perp}$$

it follows in a straightfroward way how the compute the matrices B and D. Denote the $(li - n_o) \times l$ blocks of $(\Gamma_i^{\perp})^t$ by K_i:

$$(\Gamma_i^{\perp})^t = (K_1\ K_2\ \ldots\ K_i)$$

and the $(li - n_o) \times m$ blocks of $(\Gamma_i^{\perp})^t H_t$ by Z_i:

$$(\Gamma_i^{\perp})^t H_t = (Z_1\ Z_2\ \ldots\ Z_i)$$

then it requires a tedious though straigtforward calculation to show that B and D are the solution to:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \ldots \\ Z_i \end{pmatrix} = \begin{pmatrix} K_1 & K_2 & \ldots & K_{i-1} & K_i \\ K_2 & K_3 & \ldots & K_i & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ K_i & 0 & \ldots & 0 & 0 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & \Gamma_{i-1} \end{pmatrix} \begin{pmatrix} D \\ B \end{pmatrix}$$

All matrices, except for B and D, follow directly from the SVD of the input-output block Hankel matrix. B and D themselves are the solution of a set of linear equations.

- Note that it is rather straightforward to generalize slightly the theorem for the case when the column space of the state space does not coincide with the observability subspace $(n_{oe} \leq n_o)$.

- The total least squares qualification follows from the analogy of the first step of the algorithm with the total least squares algorithm for sets of linear equations. However, the comparison stops there because the present algorithm goes on with computing estimates for A, B, C, D. Observe however, that from the lever theorem (chapter 5) and the fact that we are only interested in the short space, it follows that for instance the shift structure and hence the matrices A and C, will be consistently estimated if the noise on the data satisfies the required assumptions.

8.5 Identification via realization of a state sequence.

The identification approach to be presented in this section proceeds in several stages:

- First, a fundamental theorem is proved that shows how a state sequence can be obtained as an intersection of row spaces of block Hankel matrices of past and future inputs and outputs.

- Once a state sequence has been computed, a set of linear equations has to be solved in order to find the system matrices.

- If however one desires a reliable (consistent) identification in a *deductive* framework, some care has to be taken with respect to the solution of the problem in the so-called *short spaces.*

The main results of this approach are:

- The solution in *short* spaces, not only provides a consistent identification scheme, but moreover allows for a fast implementation, with perspectives to *on-line* computations.
- It is perfectly well feasible to interprete this identification approach in an *inspiration* approach: It will be shown, how one can define for a certain fixed *complexity* a criterion as to find the model with minimal *misfit.* The modelling approach provides, besides estimates of the linear system matrices, also *residuals*, which can be contributed to part of the data that are not compatible with a linear model of the specified complexity.
- Moreover, the oriented energy of the residuals may be used as information for a Kalman filter, in order to estimate optimally the corresponding state sequence.
- The presented approach delivers further insight into a yet unsolved problem: the combined optimality of system identification and state estimation.

8.5.1 The state as intersection of past and future.

In theorem 2, it was proved that the state vector sequence can be obtained from linear combinations of input and output block Hankel matrices. However, if only input and output measurements are available, these linear combinations, which are stated in terms of the extended observability matrix and the block Toeplitz matrix with the Markov parameters of the system, are unknown. In this section, it is shown how the state sequence can be obtained via another approach, which perfectly fits in the conceptual framework surrounding the notion of 'state of a system'.

Let's first introduce some notation: U_{h1} and Y_{h1} are block Hankel matrices as defined previously, with block dimensions i and j, containing the inputs, resp. outputs from time k to $k+i+j-2$. U_{h2} and Y_{h2} are block Hankel matrices of the same dimensions, with inputs, resp. outputs from time $k+i$ to $k+2i+j-2$. U_{h1} and U_{h2} are simply the upper- resp. the lower half of a 'big' $2mi \times j$ block Hankel matrix U_h. A similar observation holds for Y_{h1} and Y_{h2}. The data sequences contained in U_{h1} and Y_{h1} will be termed, the *past* data, while those in U_{h2} and Y_{h2} will be called the *future* data. We shall also use two block Hankel matrices, H_1 and H_2, which are simply the concatenation of input and output block Hankel matrices:

$$H_1 = \begin{pmatrix} Y_{h1} \\ U_{h1} \end{pmatrix} \qquad H_2 = \begin{pmatrix} Y_{h2} \\ U_{h2} \end{pmatrix}$$

The matrix H_1 will be referred to as the *past* block Hankel matrix, while H_2 is the *future* block Hankel matrix. The vector sequence X_1 contains the *past* states from time k to $k+j-1$ while X_2 contains the *future* states from time $k+i$ till $k+i+j-1$. Remind that a state space model is only determined up to a similarity transform.

Definition 3 Realization of a state sequence.
A vector sequence X is a realization of a state sequence X_1 if there exists a non-singular matrix T such that $X = TX_1$.

Lemma 4 *The row spaces of all possible realizations of a state sequence, coincide.*

Proof: Trivial. □

This fact is exploited in the following theorem:

Theorem 8 The state sequence as intersection of past and future.
The row space, which is the intersection of the row space of past block Hankel matrix H_1 with the row space of the future block Hankel matrix H_2, contains the row space of any realization of the future state sequence X_2. In addition, assume that:

1. $rank(H_1) = n^1_{oe} + r^1_u - r^1_{canc}$
2. $rank(H_2) = n^2_{oe} + r^2_u - r^2_{canc}$
3. $rank(H) = n_{oe} + r_u - r_{canc}$

Then, the dimension of the intersection is given by:

$$n^2_{oe} + dim(span_{row}(U_{h1}) \bigcap span_{row}(U_{h2})) + (r_{canc} - r^1_{canc} - r^2_{canc})$$

Proof: It follows from the assumed observability that:

$$X_2 = (\Gamma_i^+ \quad -\Gamma_i^+ H_t) \begin{pmatrix} Y_{h2} \\ U_{h2} \end{pmatrix}$$

which shows that X_2 is contained within the row space of the future input-output block Hankel matrix.

$$X_1 = (\Gamma_i^+ \quad -\Gamma_i^+ H_t) \begin{pmatrix} Y_{h1} \\ U_{h1} \end{pmatrix}$$

Moreover, it is straightforward to derive that

$$X_2 = A^i X_1 + (A^{i-1}B \ \ldots \ B)U_{h1}$$

Substituting the expression for X_1, then shows that X_2 can also be written as a linear combination of the row space of the past input-output block Hankel matrix. The dimension result follows from a repeated application of the Grassmann dimension theorem [15]. □

The following comments are in order:

- The theorem gives a nice interpretation of the state, being the minimal interface between past and future data.
- Observe that if the input sequence is 'wild' enough, such that there is no intersection between the row spaces of the input block Hankel matrices and there is no rank cancellation, then the dimension of the intersection equals to observable excited dimension. If there is an intersection between $span_{row}(U_{h1})$ and $span_{row}(U_{h2})$, this indicates that the concatenation of the input block Hankel matrices is rank deficient. Hence, at least part of the input can be considered as the output of an autonomous linear system and there will be rank cancellation. This means that the term $dim(span_{row}(U_{h1}) \bigcap span_{row}(U_{h2}))$ will disappear from the dimension expression.

- Depending on the method to compute the intersection, the choice of basis in the row space of the intersection will differ. However, this is immaterial since obviously only the row space of the state sequence is of importance.

- Observe that a priori no causality conditions are imposed, contrary to the least squares and the total linear least squares approach.

The following property delivers a nice possibility for the identification of systems when some of the input-output data are missing:

Corollary 5 *Let $\mathcal{K}$ be an index set with integers k satisfying $1 \leq k \leq j$. Define by H_1' and H_2' the matrices obtained from the block Hankel matrices H_1 and H_2, by omission of the columns with index in $\mathcal{K}$. Let the rows of X be a basis for the intersection between the row spaces of H_1 and H_2. Then any basis of the intersection of the row spaces of H_1' and H_2', delivers a state sequence X', which is a realization of the state sequence X, but with omitted states with indices in $\mathcal{K}$.*

Proof: Immediate from theorem 7. □

A immediate consequence of this lemma is, that *the block Hankel structure is not necessary in our identification approach.* However, observe that the (block) Hankel structure allows for an economic exploitation of the number of data.

Observe that corollary 5 allows to analyse the presented identification approach from a deductive point of view, using the lever and the orthogonality theorems of chapter 5, when the assumptions on the additive noise are less restrictive. Indeed, the block Hankel structure requires the assumption for the noise to be white, while in the absence of any structure, this assumption may be relaxed.

8.5.2 Computing an approximate state realization.

Having stated the possibility of computing a state sequence from input-output data only in theorem 7, it will now be shown how this can be done apporoximately when the data are noisy.

While canonical correlation analysis was developped independently by Hotelling and Obukhov in the thirties, its potentiality for use in dynamical system identification and data analysis was only discovered recently. Yaglom and Gelfand observed that the past and the future output processes of a stochastic process have a finite number of non-zero canonical correlations if and only if the process has a rational power spectrum, i.e. it is a finite order Markov process [27]. This observation delivers a meaningful interpretation of the intersection of the past and future block Hankel matrices. Also in [27], one may find a nice literature survey on the field. Akaike has developed a stochastic realization theory based on the information interface between the past and the future of a time series [2] [3]. Extensions of this work were derived by Baram, Desai and Pal, Verriest [36] [45] [46]. In general, all cited authors have derived their results in a stochastic framework, *explicitly assuming that the input of the system, is an unknown but white stochastic process.* This assumption facilitates considerably the computation and interpretation of the results. In our approach however, no specific assumptions on the input

sequence is made.

The problem formulation reduces to the computation of an approximate state sequence X, which reduces to the determination of an appropriate approximate intersection. Hereto, the results derived in chapter 5, will be used. While in chapter 5, these results were developed mainly based on mathematics, we shall now add an extra dimensions by interpreting them from the conceptual point of view, in the context of dynamical state space descriptions. We shall describe how to compute an approximate state space sequence for the following cases, the main difference between them being the approximation criterion.

1. Least Squares Approach for noisy outputs, canonical correlations.
2. Total Linear Least Squares Approach, canonical correlations.
3. Forward and Backward Predictor approach, canonical correlations.
4. Optimal predicition, unnormalized computation of the intersection.

Least squares for noisy outputs.

In some applications, the inputs are known more accurately than the outputs. This case can easily be dealt with for the computation of an approximate intersection. Denoting by $U_{h1}, Y_{h1}, U_{h2}, Y_{h2}$ past and future input and output block Hankel matrices, an approximate state sequence can be found from the solution of:

$$min_{P_1,Q_1,P_2,Q_2} \ \|(P_1^t U_{h1} - P_1^t U_{h2}) - (Q_1^t Y_{h1} + Q_2^t Y_{h2}\|_F^2$$

subject to the constraint:

$$XX^t = \Delta$$

where X represents the obtained state sequence and Δ is any positive definite diagonal matrix. The solution is a simple application of theorem 13 of chapter 5.

Total Linear Least Squares Approach.

In case the inputs and the outputs of the system are corrupted by comparable amounts of noise (possibly after scaling), the following criterion is the most natural:

$$min_{X,P,Q} \ \|X - P^t H_1\|_F^2 + \|X - Q^t H_2\|_F^2$$

subject to the constraint:

$$XX^t = \Delta$$

where X is the state sequence and Δ is diagonal positive definite. The solution follows immediately from a simple application of theorem 14 of chapter 5. As a most natural fact, it follows for instance that:

$$X = \frac{1}{2}(P^t H_1 + Q^t H_2)$$

P and Q can be computed from the generalized singular value decomposition of the matrix pair $(H_1 \ H_2)$ as explained in theorem 14 of chapter 5.

Forward and Backward predictor subspaces.

The past block Hankel matrix H_1 is considered to contain all relevant information about the past. The new information in H_2 is contained in that subspace of the row space of H_2 which is *orthogonal* with respect to the row space of H_1. This looks very much alike the idea and notion of *innovations* and it will now be shown that its solution reduces to solving a linear least squares problem. The criterion to be optimized is:

$$min_{P,Q} \ \|P^t H_1 - Q^t H_2\|_F^2$$

subject to:

$$XX^t = \Delta$$

where the state sequence X equals $X = Q^t H_2 H_1^+ H_1$ and Δ is a positive definite diagonal matrix, of dimensions $n \times n$. n is the proposed dimension of the observable excited subspace of the state space and may be derived, for instance by inspection of the canonical angles. The solution follows immediately from the result summarized in theorem 14 of chapter 5.

The following comments are in order:

- The essential point about the constraint on the state sequence X is the fact that Δ is of full rank n. From the computational point of view, it is completely unimportant that Δ is diagonal since when it were not, any diagonalization transformation would deliver an equivalent state realization, which simply corresponds to a change of basis in the state space. The diagonality requirement however simplifies the computation of the solution, as derived in chapter 5.

- Although further investigation is needed, we conjecture that it must even be possible to chose a priori an appropriate matrix Δ, positive definite but not necessarily diagonal, such that for instance the obtained model is at once in balanced state space coordinates. Promising results in this direction have recently be obtained in this direction [45].

- From the expression for the state sequence X and the conceptual interpretation of the least squares identification scheme, it follows that the row space of the approximate state sequence X is a subspace of the row space of H_1. This is the mathematical translation of the assumption, that the past contains all necessary information about the future or stated in other terms, the information that is present in the past, is used to predict as good as possible, the future behavior.

Not in all applications, the lack of an intersection between the row spaces is due to noise. For instance, when the system is time varying or non-linear, the presented identification approach may be applied from a classical quasi-stationary assumption, using a gliding window approach with or without adaptive weighting, in the hope that the identified *time invariant linear* system provides a good *approximation* of the system, at least within the window. Two appropriate state sequences can be computed using the RV-coefficient approach presented in theorem 15 of chapter 5. One of them is an approximate state sequence of which the row space belongs to the row space of the past block Hankel matrix. The other one is a state sequence of which the row space belongs to the row space of the future block Hankel matrix.

For obvious reasons, they are called the *forward predictive* and the *backward predicitive* state realizations. The proposed criterion is the following:

$$max_{P,Q} \;\; RV(P^t H_1 \; , \; Q^t H_2)$$

subject to the constraints:

$$P^t H_1 H_1^t P = \Delta_{H_1} \qquad Q^t H_2 H_2^t Q = \Delta_{H_2}$$

where Δ_{H_1} and Δ_{H_2} are diagonal positive definite matrices. The solution for the matrices P and Q follows from the generalized singular value decomposition of the matrix pair $(H_1 \; , \; H_2)$ as was descibed in section 5.4.9. The forward and backward predictor state sequence are then given by:

$$X_{forward} = P^t H_1 \qquad X_{backward} = Q^t H_2$$

Observe that the constraints do *not* allow to ensure that the row space of the approximation, represented by the average of the backward and the forward state $(X_{forward} + X_{backward})/2$, has the required dimension n, although *generically*, this will be the case. However, it is precisely in this respect that the forward-backward prediction approach differs from the total linear least squares approach, though both of them are based upon canonical correlation analysis.

Optimal predicition, unnormalized.

Recall from chapter 5, that the canonical correlation coefficients are not only biased when the data are inexact, but that moreover they can be very misleading in the determination of the system order. The reason is the inherent *normalization* to orthonormal matrices when computing the canonical correlation structure. Hence one empasses completely the energy information concerning the relative strength of the different modes that are present in the measurements. When employing canonical correlation analysis, high energetic and low energetic modes are treated on the same footing. A better, *unnormalizing* approach was proposed in section 5.4.10. When applied to dynamic system identification, one finds the following results:
An approximate *forward predictor* state sequence can be found from optimization of the criterion:

$$min_{P,Q} \;\; \|Q^t H_2 - Q^t H_2 H_1^t P (P^t H_1 H_1^t P)^{-1} P^t H_1 H_2^t Q\|_F^2$$

subject to the constraints:

$$P^t H_1 H_1^t P = \Delta_{H_1} \qquad Q^t H_2 H_2^t Q = \Delta_{H_2}$$

where Δ_{H_1} and Δ_{H_2} are positive definite matrices that ensure that the approximate forward predictor state sequence, $X_{forward} = P^t H_1$ has the required rank. As was demonstrated in theorem 16 of chapter 5, the solution is obtained from the generalized singular value decomposition of the matrix pair $(H_1 \; , \; H_1 H_2^t)$. The generalized singular values allow for a meaningful estimation of the dimension of the state space, taking into account the relative energy of the several modes present in the signal.

8.5.3 Solving for the system matrices.

Exact solution

Once a realization of the state sequence X_2 is known, the system matrices A, B, C, D are easily solved from a set of linear equations:

$$\begin{pmatrix} x_{k+i+1} & \cdots & x_{k+i+j-1} \\ y_{k+i} & \cdots & y_{k+i+j-2} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_{k+i} & \cdots & x_{k+i+j-2} \\ u_{k+i} & \cdots & u_{k+i+j-2} \end{pmatrix}$$

If the data are exactly generated by a linear system, then these linear equations will be exactly solvable and provide a realization of the system matrices A, B, C, D. However, as could be expected the scene changes considerably whenever the data are corrupted by noise or when only an approximate state sequence is available.

Approximate solutions.

First note that, the fact that the system of linear equations in the unknown matrices A, B, C and D does not posses an exact solution, might have several causes: The data may be perturbed by additive (measurement) noise, they may be generated by a non-linear system, they may be generated by a linear system but with unobserved inputs, the state sequence may have been determined only approximative.

In any case, the set of linear equations can only be solved in an approximate way if the data are not exact. This results in *residuals* that contain the misfit. The residuals of the solution can later on be exploited in for instance a Kalman filter.
Assume that the computed approximate state sequence consists of the vectors x_k and that the observed inputs and outputs are given by y_k and u_k. Several solution strategies may be applied in order to solve the set of linear equations. For instance, if it is known or assumed that the inputs are noise free while the states and outputs are not, one should try to estimate the relative noise levels on the states in the outputs, scale the corresponding data appropriately in order to bring them on the same noise level, then apply a RQ factorization in order to decouple the exact inputs from the noisy states and outputs and obtain the solution via the mixed least squares - total linear least squares approach of lemma 1 of section 5.3.2. If the inputs are noisy and the outputs noise free, simply apply a similar strategy. If all data are noisy, *scaling* is necessary in order to make the noise level the same for all data, whereafter one can apply a 'full' total least squares strategy.

Whatever method be used to solve the set of linear equations, the derived solution can be considered as the exact solution to a set of linear equations with 'corrected' data of the from:

$$\begin{pmatrix} x^{cc}_{k+1} \\ y^{c}_{k} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x^{c}_{k} \\ u^{c}_{k} \end{pmatrix}$$

The residuals can be computed as:

$$\begin{aligned} \tilde{x}^{cc}_{k+1} &= x_{k+1} - x^{cc}_{k+1} \\ \tilde{x}^{c}_{k} &= x_{k} - x^{c}_{k} \\ \tilde{u}_{k} &= u_{k} - u^{c}_{k} \end{aligned}$$

$$\tilde{y}_k = y_k - y_k^c$$

Observe that the residual on the state x_{k+1} has no relation to the residual on the state x_k. The reason is that the mathematical solution of a set of linear equations, taking explicitly into account the structure between the rows with data generated by x_{k+1} and those, generated by x_k, is *a highly complicated and unsolved* problem. Thereto, we adopt a more pragmatic point of view by not taking into account the structure, but adding residual terms to the identified model, that can be interpreted as process and measurememt noise. Several choices are possible, depending on the specification of the approximate state:

State $= x_k^c$**:**

$$\begin{pmatrix} x_{k+1}^c \\ y_k \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_k^c \\ u_k \end{pmatrix} + \begin{pmatrix} x_{k+1}^c - x_{k+1}^{cc} - B\tilde{u}_k \\ \tilde{y}_k - D\tilde{u}_k \end{pmatrix}$$

State $= x_k^{cc}$**:**

$$\begin{pmatrix} x_{k+1}^{cc} \\ y_k \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_k^{cc} \\ u_k \end{pmatrix} + \begin{pmatrix} A(x_k^c - x_k^{cc}) - B\tilde{u}_k \\ \tilde{y}_k - D\tilde{u}_k + C(x_k^c - x_k^{cc}) \end{pmatrix}$$

State $= x_k$**:**

$$\begin{pmatrix} x_{k+1} \\ y_k \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_k \\ u_k \end{pmatrix} + \begin{pmatrix} \tilde{x}_{k+1}^{cc} - A\tilde{x}_k^c - B\tilde{u}_k \\ \tilde{y}_k - D\tilde{u}_k - C\tilde{x}_k^c \end{pmatrix}$$

Observe that the proposed models differ only in the definition of the state which determines the residuals. The first choice for the state delivers the most simple model for the residuals in this sense, that it is not determined directly by the model matrices. This choice will be appropriate when the data originate from a linear model, when the process noise is white and the measurement noise is Gaussian. Of course, if these stochastic characterizations are unknown a priori (as is the case in most practical situations), a rigorous qualification of the obtained model is much more difficult to assess. Further (conceptual) research is needed in order to motivate rigorously the choice of the proposed state sequence and the associated optimality properties. Certainly, this problem is not new since in his publication of 1960, Kalman [24] observes that:

" ...As in the vast majority of the engineering literature on the Wiener problem, we shall find it convenient to start with the model and regard the problem of obtaining the model as a separate question. To be sure, the two problems should be optimized jointly if possible ; the author is not aware however, of any study of the joint optimization problem ".

The oriented energy of the residuals is the necessary information for a succesful application of the Kalman filter. It is not too difficult to see that the Grammians of the residual sequences are sample estimates of the necessary covariance matrices needed in the Kalman filter, that act as pre-whitening (Mahalanobis) transformations as described in chapter 4. If it could be proved that the derived modelling strategy is optimal in a certain sense, then also the

combined approach identification-state estimation could further be analysed for its joint optimality. However, recall from our deductive analysis in chapter 5, that the state sequence, as approximated in the *long* space, is *an inconsistent estimate.* However, since the model matrices are derived from the *short* space, the model can be proved to be consistent.

8.5.4 Reconversion to the short space.

One of the fundamental conclusions of chapter 5, was that the matrices in the *short* space, that describe the linear combinations which provide the 'long' vectors of the intersection, may be estimated *consistently*, but *not* the 'long' vectors themselves. Since the rows of the state sequence are 'long' vectors from the intersection of the row spaces of H_1 and H_2, this seems to suggest that it is difficult (if not impossible) to estimate the state sequence itself *consistently*, directly from input-output measurements, hence suggesting a two stage identification approach. It will now be shown that all necessary information can be recovered from the *short* spaces of H_1 and H_2.

1. First, the *short* spaces provide a kind of *dual* model for the intersection. It will be shown that the system matrices are identifiable, using only information obtained from the *short* spaces.

2. The very fact that everything is formulated in the short space, leads to an efficient computational implementation, and even to a fast adaptive algorithm.

The derivation will be presented for 'exact' data, as is always the case in a *deductive* approach. When the data are noisy, the perturbation results obtained in chapter 5, carry over without modification.

In order to avoid unnecessary complicated notation and the detailed analysis of special cases (the treatment of which deviates in no way of the general presentation), we shall assume from now on that the following (plausible) assumptions hold true:

- There is no rank cancellation.

- The matrices U_{h1} and U_{h2} are of full row rank $rank(U_{h1}) = rank(U_{h2}) = mi$.

We shall discuss two possible approaches for the reduction of the algorithms to the short spaces: one exploits the idea based upon the 'economy size' SVD of corollary 8 of section 5.4.3. This is the heuristic approach. The other one is based upon the QR-factorization idea, developed in section 5.4.11. which allows for an integration in the optimization criteria of section 5.4.
Both techniques allow to reduce the number of flops considerably.

Economy size intersection computation.

Denote the concatenation of H_1 and H_2 by H:

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$$

and assume that its singular value decomposition is given as:

$$H = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_{11} & 0 \\ 0 & 0 \end{pmatrix} V^t$$

Assuming the the observable excited dimensions equals n, these matrices have the following dimensions:

$$\begin{aligned} U_{11} &= (mi + li) \times (2mi + n) \\ U_{12} &= (mi + li) \times (2li - n) \\ U_{21} &= (mi + li) \times (2mi + n) \\ U_{22} &= (mi + li) \times (2li - n) \\ S_{11} &= (2mi + n) \times (2mi + n) \end{aligned}$$

Now from:

$$(U_{12}^t \; U_{22}^t) \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = 0$$

or equivalently:

$$U_{12}^t H_1 = -U_{22}^t H_2$$

it follows that the row space of $U_{12}^t H_1$ equals the required intersection of the row spaces. However, the matrix $U_{12}^t H_1$ contains $2li - n$ row vectors, of which only n are linearly independent according to theorem 2. Hence it remains to select n suitable linear combinations of these $2li - n$ rows. One straightforward way would consist in computing the singular value decomposition of $U_{12}^t H_1$ in order to compute a basis for the row space. However, this matrix is a $(2li - n) \times j$ matrix, where j is usually very large. The following theorem provides a shortcut to the computation of this large SVD, by replacing it by a smaller SVD:

Theorem 9 *Let the SVD of the concatenation of past and future block Hankel matrix be given by:*

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_{11} & 0 \\ 0 & 0 \end{pmatrix} V^t$$

then a realization of the state sequence X_2 can be calculated as:

$$X_2 = U_q^t U_{12}^t H_1$$

where U_q is an $(2li - n) \times n$ orthonormal matrix defined through the SVD of:

$$U_{12}^t U_{11} S_{11} = (U_q \; U_q^\perp) \begin{pmatrix} S_q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_q^t \\ (V_q^\perp)^t \end{pmatrix}$$

Proof: Since any basis for the row space of $U_{12}^t H_1$ is a realization of the state sequence X_2, we first calculate its SVD:

$$\begin{aligned} U_{12}^t H_1 &= U_{12}^t (U_{11} \; U_{12}) \begin{pmatrix} S_{11} & 0 \\ 0 & 0 \end{pmatrix} V^t \\ &= (U_{12}^t U_{11} S_{11} \;\; 0) V^t \\ &= (U_q S_q V_q^t \;\; 0) V^t \\ &= U_q (S_q \;\; 0)(V V_q)^t \end{aligned}$$

Hence we derive easily that:

$$U_q^t U_{12}^t H_1 = (S_q \;\; 0) V_q^t V^t$$

which is a valid basis for the row space of $U_{12}^t H_1$ and hence must be a realization of the state sequence X_2. □

QR-decomposition of the matrix H.

Most of the identification algorithms based upon the computation of an approximate intersection, resulted in a solution of the form:

$$min_{P,Q} \| P^t H_1 - Q^t H_2 \|_F^2$$

and the optimizing matrices P and Q could be derived from a generalized singular value decomposition of H_1 and H_2 as described in section 5.4. In section 5.4.11, it was shown how a preliminary QR-factorization of the concatenated matrix could reduce considerably the required amount of flops, without loss of accuracy:

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} R_1 & 0 \\ R_2 & R_3 \end{pmatrix} Q$$

All necessary information is now concentrated in the left factor R and one can compute the required GSVD from this matrix only.

Solving for the model matrices in the short space.

A most interesting observation is that, whatever economizing preprocessing factorization is applied (one of the two described), it carries through to the solution of the set of linear equations, which is the second step of an identification algorithm that first realizes an approximate state. We shall demonstrate this result for the case of the 'economy size' SVD approach, but first we need to introduce some (PC-Matlab) matrix notation:

- $M(p:q, r:s)$ is the submatrix of M formed by rows p to q and columns r to s.
- $M(:, r:s)$ is the submatrix of M generated by columns $r, r+1, \ldots, s$.
- $M(p:q, :)$ is the submatrix of M containing rows $p, p+1, \ldots, q$.

We shall also use slightly reorganized matrices H_1 and H_2:

$$H_1 = \begin{pmatrix} u_k & u_{k+1} & \cdots & u_{k+j-1} \\ y_k & y_{k+1} & \cdots & y_{k+j-1} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+j} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+j} \\ \cdots & \cdots & \cdots & \cdots \\ u_{k+i-1} & u_{k+i} & \cdots & u_{k+i+j-2} \\ y_{k+i_1} & y_{k+i} & \cdots & y_{k+i+j-2} \end{pmatrix}$$

$$H_2 = \begin{pmatrix} u_{k+i} & u_{k+i+1} & \cdots & u_{k+i+j-1} \\ y_{k+i} & y_{k+i+1} & \cdots & y_{k+i+j-1} \\ u_{k+i+1} & u_{k+i+2} & \cdots & u_{k+i+j} \\ y_{k+i+1} & y_{k+i+2} & \cdots & y_{k+i+j} \\ \cdots & \cdots & \cdots & \cdots \\ u_{k+2i-1} & u_{k+2i} & \cdots & u_{k+2i+j-2} \\ y_{k+2i-1} & y_{k+2i} & \cdots & y_{k+2i+j-2} \end{pmatrix}$$

Let the SVD of the concatenation H of H_1 and H_2 be given by:

$$H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = USV^t = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_{11} & 0 \\ 0 & 0 \end{pmatrix} V^t$$

and the SVD of the matrix $U_{12}^t U_{11} S_{11}$ as:

$$U_{12}^t U_{11} S_{11} = (U_q \; U_q^\perp) \begin{pmatrix} S_q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_q^t \\ (V_q^\perp)^t \end{pmatrix}$$

then we have that:

Theorem 10 Reduced Set of Linear Equations
The system matrices A, B, C, D can be identified from the set of linear equations:

$$\begin{pmatrix} U_q^t U_{12}^t U((m+l+1):((i+1)(m+l)),:)S \\ U((mi+li+m+1):((m+l)(i+1)),:)S \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} U_q^t U_{12}^t U(1:(mi+li),:)S \\ U((mi+li+1):(mi+li+m),:)S \end{pmatrix}$$

where the notations and dimensions are the same as in theorem 9.

Proof: From theorem 9, we find immediately that:

$$\begin{aligned} (x_{k+i} \; \cdots \; x_{k+i+j-1}) &= U_q^t U_{12}^t H_1 \\ &= U_q^t U_{12}^t H(1:(mi+li),:) \\ &= U_q^t U_{12}^t U(1:(mi+li),:)SV^t \end{aligned}$$

From the structure of the matrices, it is also easily seen that:

$$\begin{aligned} (x_{k+i+1} \; \cdots \; x_{k+i+j}) &= U_q^t U_{12}^t H((m+l+1):((i+1)(m+l)),:) \\ &= U_q^t U_{12}^t U((m+l+1):((i+1)(m+l)),:)SV^t \end{aligned}$$

In the same way, it is found that:

$$\begin{aligned} (u_{k+i} \; \cdots \; u_{k+i+j-1}) &= H((mi+li+1):(mi+li+m),:) \\ &= U((mi+li+1):(mi+li+m),:)SV^t \\ (y_{k+i} \; \cdots \; y_{k+i+j-1}) &= H((mi+li+m+1):((m+l)(i+1)),:) \\ &= U((mi+li+m+1):((m+l)(i+1),:)SV^t \end{aligned}$$

The result follows immediately by substitution of the derived block vectors into the equation:

$$\begin{pmatrix} x_{k+i+1} & \cdots & x_{k+i+j}) \\ y_{k+i} & \cdots & y_{k+i+j-1} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_{k+i} & \cdots & x_{k+i+j-1} \\ u_{k+i} & \cdots & u_{k+i+j-1} \end{pmatrix}$$

and postmultiplication with V then proves the theorem. □

Observe that:

- We have derived the theorem assuming exact data. Of course, in the case of noisy data, one applies the algorithm formally, using approximations of reduced rank where necessary.
- The reduction of the computational complexity arises from discarding the common orthogonal factor V^t which is a 'very rectangular' matrix. Observe that the same result would apply, without much modification, if first a RQ-factorization of the concatenated matrix H would have been performed. The common factor Q could also be discarded without any complication.
- From the lever theorem in chapter 5, it follows that the discarded orthonormal factor, being an orthonormal basis for the *long* space, is the most sensitive to additive noise. The observation that it does not contribute to the estimation of the model matrices, confirms the statement that the model matrices are estimated *consistently*, since all the necessary information is contained within the short space.
- Observe that omitting also the diagonal matrix S containing the singular values, corresponds to a particular choice of weighting the set of linear equations.
- As a matter of fact, since the discarded factors are orthonormal, it is straightforward to show that there is no difference in the (total) least squares solutions of the *long* set of equations compared to those of the *short* sets. This is an important observation if in addition to the identification algorithm, also a Kalman filter is applied in order to estimate optimally the state sequence.

Corollary 6 The residuals.
Let the set of linear equations to be solved for the matrices A, B, C, D be reduced to a 'shorter' set by discarding a common orthogonal factor. Then the Grammians of the (total) linear least squares residuals are the same in both cases.

Proof: This follows immediately from the *orthonormality* of the discarded common factor. □

8.5.5 An on-line algorithm

Adaptive estimation may be required for several reasons:

- Some of the characteristics of the system may be intrinsically uncertain or fuzzy, hence necessitating an adaptive strategy to estimate an appropriate model.

- The system may be non-linear but it is approximated by a linear one within a window of finite length. This linearisation approach can be applied whenever the absolute time duration of the finite length of the window is sufficiently small with respect to the sampling time. The misfit, due to the difference between the approximate linear model and the non-linear is then contributed to the residuals. However, while intuitively this approach seems very plausible, its theoretical and conceptual analysis is highly non-trivial.

- Time variance of the system may be dealt with in a similar approach, i.e. considering the system as time invariant in a sufficiently small finite window. Some insight into the timevariance can be obtained from the update of the model in the gliding window. Typically, a linear time variant system can be approximately modelled (especially in short windows) by a *linear time invariant* model, but of a higher complexity.

As an example of the complexity of the theoretical foundations of adaptive state space equations, consider the problem of updating the state space basis i.e. the number of linearly independent vectors in relation to which are expressed the consecutive states obtained in the computation of an approximate intersection. When it is known a priori that the system is time invariant and the data originate in a linear system without any noise, the solution is straightforward. Assume that two state sequences have been obtained in two overlapping windows. Then the solution reduces to the computation of a matrix that represents the linear transformation between the two state space bases. This can be done from the solution of a set of linear equations using state realizations of both windows from the overlapping time instants. However, the solution ceases to be easy when for instance the system is timevarying or non-linear or when the data are noisy. In these cases, there does not exist a simple relation between the state realizations of different, but overlapping time windows. Of course, one could define some measure which allows to compute an approximation of the (non-linear, time varying, ...) transformation, as for instance some linear least squares criterion. Such a criterion is obviously inspired by the easyness of its solution. It is extremely difficult to motivate rigorously the choice of such criterion from a conceptual point of view.

There are two possible remedies:

1. The Kalman filter provides an *indirect* approach since its formulation allows explicitly for time-varying system matrices. Hence one finds the optimal state taking into account a possible change of basis in the state space. Of course, this is not completely satisfactory: Assume that a time-invariant state feedback control strategy is applied and that the real system is really time-invariant linear. Assume that a gliding window adaptive identification approach delivers the correct model, but that in every window this is expressed with reference to another basis. Then the Kalman filter will deliver in every window the optimal state but it is difficult to apply constant state feedback, without updating the state space basis.

2. At least for time varying systems, a more conceptually oriented way out is maybe provided by some recent results on balanced state space models for time varying systems [45]. The noisy case however remains largely *terra incognita*. Clearly, this represents a serious drawback of any state space approach with respect to an input-output approach where at first instance, no direct estimate of the state is needed.

Another serious problem is the adaptive estimation of an appropriate observable excited dimension. In our opinion, the adaptive estimation of the order is an an essential feature of an adaptive identification approach. In most identification strategies, fixing the complexity of the model, which in most cases is proportional to the number of parameters that is to be estimated, leads to ill conditioning of the algorithm and associated numerical instabilities whenever the fixed complexity exceeds the dimensionality of the excited dynamics. This is especially the case in an adaptive control environment. If the control action aims at eliminating dynamic modes of the system, the excited observable dimension will decrease in time when the controller performs his task sufficiently well. If however the assumed model order is not decreased as well, some steps of the (any?) identification algorithm become ill-conditioned: For instance, in trying to solve the model matrices from the estimated state sequence and the input and output observations, the least squares algorithm will deliver numerically unreliable results, because some of the matrices involved in the solution, become (almost) rank deficient. It is however easily seen that adaptive estimation of the order or complexity of the model, reduces to the ability of adaptive identifying an appropiate lower rank approximation of certain matrices. This is however not at all a trivial task. Despite the fact that several rank determination criteria have been proposed recently (Akaike's Information Criterion [4], Rissanen's Minimum Description Length [37]), much work remains to be done in order to deliver satisfactory answers to this difficult problem. In this work, it is assumed that some rule is applied that reduces essentially to a comparison of either canonical angles or (generalized) singular values with respect to some a priori fixed or motivated threshold, for instance based on our oriented signal-to-signal ratio framework of chapter 4.

Assuming that a certain complexity has been fixed, it is straightforward to outline a possible adaptive scheme for the identification of a state space model. Every time step a new input-output vector pair becomes available, a new column is to be added to the input-output block Hankel matrix H. On the other hand, when for instance the system is time-varying, older measurements should be discarded. This can be done either via a deletion of columns or by an exponential weighting approach. In this last case, instead of deleting columns, all columns are multiplied by a scaler $\alpha < 1$. Hence, a column which was added t steps earlier, is weighted with a vector α^t, thus effectively reducing its relative contribution with respect to more recently obtained and added measurements. Let's now derive an adaptive version of the economy size approach described in section 8.5.4.

Initialize $U_0 = I_{(2mi+2li)}$, $S_0 = 0$, m and l being the number of inputs and outputs, $2i$ being the number of block rows in the matrix H (that is however *never* to be constructed explicitly!).

For $k = 1, 2, \ldots$

1. Construct a new column, denoted by *column*, to be inserted to H, using the latest $2i$ I/O-measurements.

2. Update the singular value decomposition:

$$U_k S_k V_k^t = (\alpha U_{k-1} S_{k-1} \;\; column)$$

3. Invoke a *rank estimation procedure* and partition accordingly the SVD:

$$U_k S_k = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_{11} & 0 \\ 0 & S_{22} \end{pmatrix}$$

4. Compute the SVD of $U_{12}^t U_{11} S_{11}$ and partition according to the estimated rank:

$$U_{12}^t U_{11} S_{11} = (U_q \;\; U_q^\perp) \begin{pmatrix} S_{q1} & 0 \\ 0 & S_{q2} \end{pmatrix} \begin{pmatrix} V_q^t \\ V_q^{\perp t} \end{pmatrix}$$

5. Solve the set of linear equations:

$$\begin{pmatrix} U_q^t U_{12}^t U((m+l+1):((i+1)(m+l)),:)S \\ U((mi+li+m+1):((m+l)(i+1)),:)S \end{pmatrix}$$
$$= \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} U_q^t U_{12}^t U(1:(mi+li),:)S \\ U(((mi+li+1):(mi+li+m)),:)S \end{pmatrix}$$

end

The following remarks apply:

- The adaptive solution of the set of linear equations could be performed via a 'classical' recursive linear least squares solver, taking into account only one additional column at each time step. This would result in an additional speeding up of the algorithm. Observe that the knowledge of the residuals in each step can be updated in order to arrive at optimal *prewhitening* (Gauss-Markov) [34] estimation algorithms. Moreover, the required algorithms allow for a *fast* implementation.
- In order to derive adaptive algorithms for the criterion optimizing approaches described in section 8.5.2., an in depth research is required for rank-one or rank-two update strategies of the *generalized* singular or eigenvalue- decomposition. The necessity for this originates in the rank constraints imposed on the approximate intersection. This forces a Lagrange multiplier approach resulting in generalized eigenvalue approaches.
- Observe however that also the heuristic approach requires rank-one update algorithms for the singular value decomposition.
- A largely unexplored domain of research is an explicit exploitation of the (block) Hankel structure of the matrices, taking into account the required numerical stability. A suggestion may be contained in some recent results on *displacement rank concepts* and related algorithms.

8.6 Examples and applications

8.6.1 A power plant

Input-output measurements on a 120 MW power plant of Pont-Sur-Sambre in France were obtained. The available input data were samples under *normal operating records* of the gas

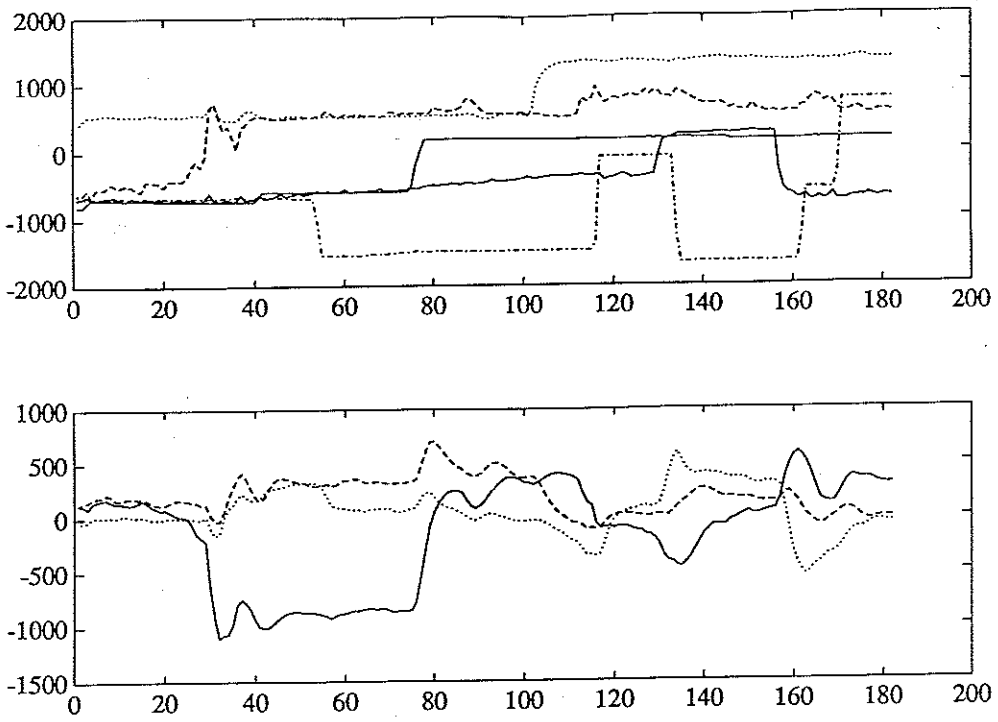


Figure 8.11: 5 inputs and 3 outputs of a power plant.

flow, the turbine valves opening, the super heater spray flow, the gas dampers and the air flow. The three outputs are the steam pressure, the main steam temperature and the reheat steam temperature. The sampling time was 120 seconds. Details about the data acquisition can be found in [20].

Both the heuristic total least squares approach (section 8.4.) and the economy size algorithm (section 8.5.5.) were applied and provided excellent models. The results on the total least squares approach can be found in [42]. Here we shall demonstrate how the economy size algorithm appears to be very robust with respect to the order estimation. From figure 8.12, it can be seen that the quality of the simulation ameliorates, as the complexity of the model is increased. The block dimensions used for the canonical correlation approach were $i = 5, j = 90$.

8.6.2 An ecological system

Data regarding the lake Erie in the period March 1968 - November 1972 were collected and identified. Details concerning data-acquisition can be found in [21]. The five inputs are the water temperature, the water conductivity, the water alkalinity, the dissolved NO_3 and the total hardness. The two outputs are given by measures of the algae present in the water and the dissolved oxigen. The applied identification method is the total linear least squares approach described in theorem 7 with as block dimensions for the input-output block Hankel matrix $i = 4, j = 54$. The results can be found in figure 8.13. The depicted singular spectrum only shows the singular values from the 21-th on. The first 20 singular values originate from the input block Hankel matrix. It can be seen that the order determination is not always a trivial task. The order was here chosen as $n = 4$.

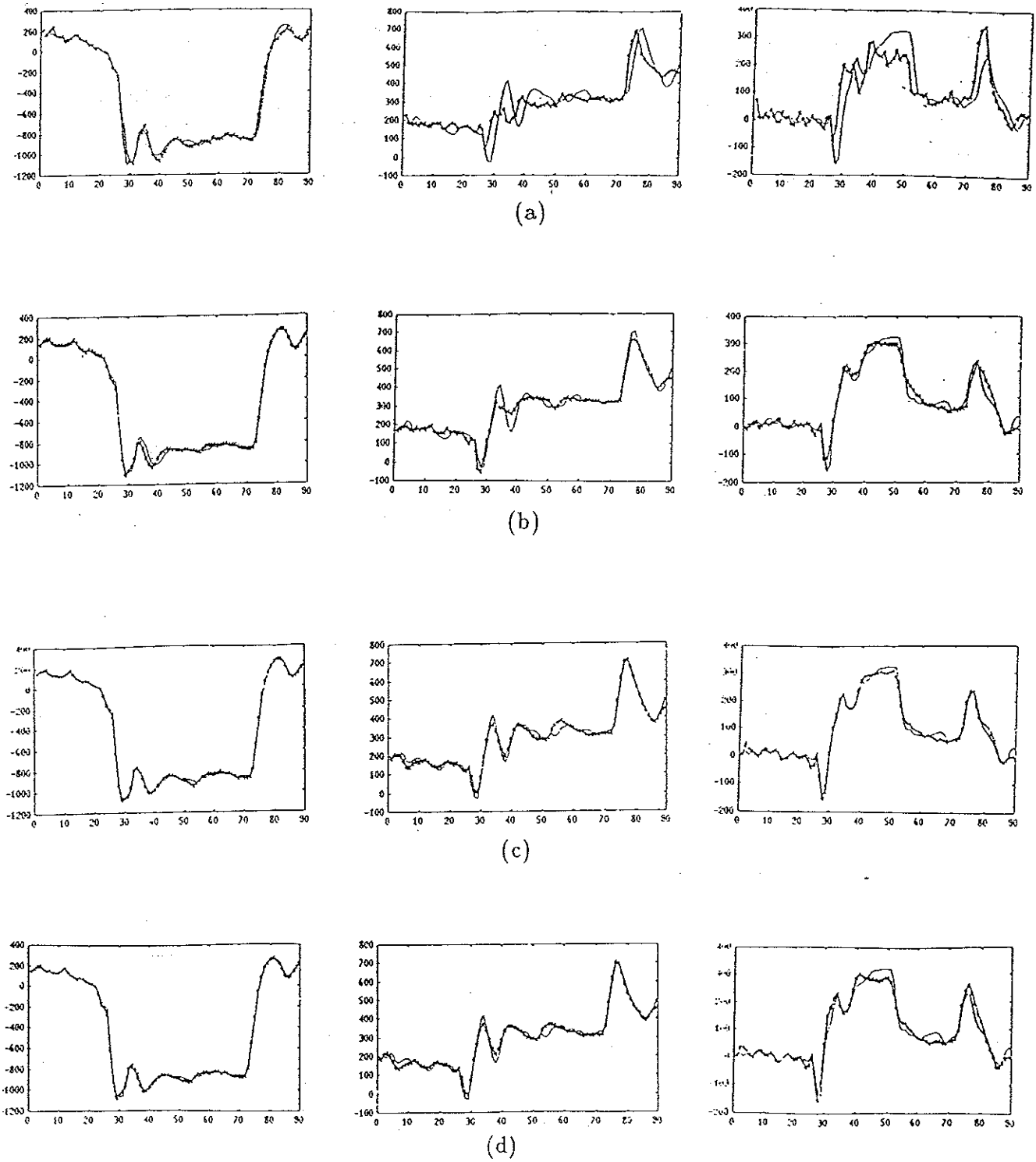


Figure 8.12: Measured outputs (full line) and simulations (stars) for a (a) first order, (b) 4-th order, (c) 7-th order, (d) 9-th order model

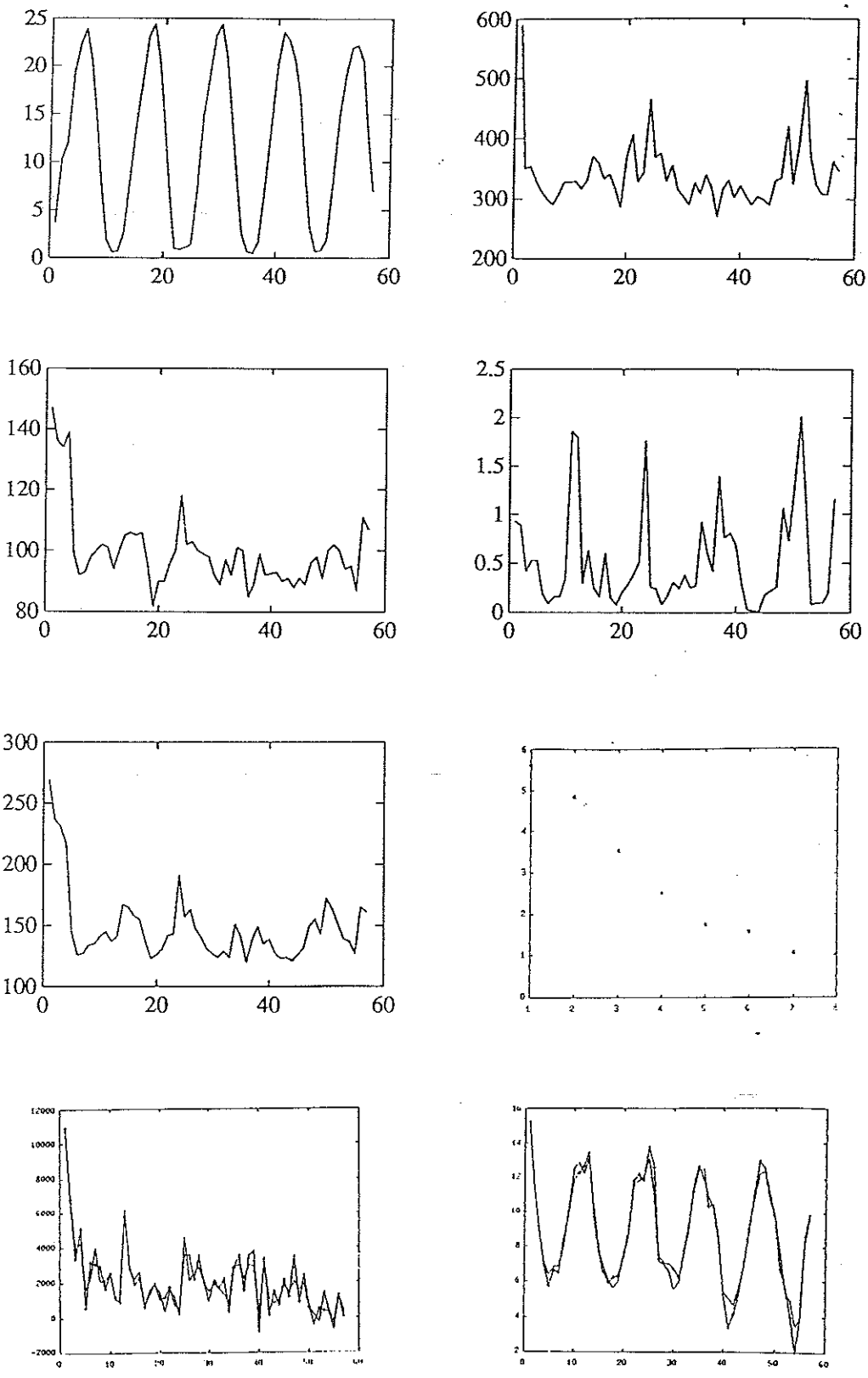


Figure 8.13: 5 inputs, singular spectrum of input-output block Hankel matrix, 2 outputs, measured(full), simulated (+). Time in months.

8.6.3 A chemical destillation column.

Measurements on an ethane-ethylene distillation column were collected [21]. The five inputs of the model are the ratio between the top flow and the feed, the ratio between the reflux ratio and the feed, the ratio between the distilled output and the feed, the ethane feed and the top pressure. The three outputs are measurements on the top ethane flow, the bottom ethylene flow and the differential pressure between the top and the bottom. The sampling time is 15 minutes. Other details concerning the data-acquisition can be found in [21]. The identification method is the 'economy size' approach, analysed in theorem 10. The block dimensions of the block Hankel matrices are $i = 3, j = 85$ and the order is determined as $n = 6$. The simulations are shown in figure 8.14.

8.6.4 A glass production installation

A feeder is the final part of a process installation that is used for melting glass, in which the glass is conditioned for further processing. Its mean task is to realize a homogeneous temperature distribution. The reader is invited to consult [8, p.193] for an in depth discussion of the data-acquisition and signal preparation. Pseudo-random binary sequences have been applied simultaneously to three process inputs: Input 1 is a gas input of the first feeder section, input 2 is a cooling air input while input 3 is the gas input of the second feeder section. The first 300 samples of these inputs are depicted in figure 8.15. As outputs of the process, the glass temperature at 6 points in a cross section of the feeder close to the spout have been measured (see [8, p.202]). The identification method that was used is the canonical correlation method, described in theorem 9 and 10. The block dimensions of the input-output block Hankel matrices are $i = 10, j = 300$. The singular spectrum of the $2(m+l)i \times j$ (180×300) is depicted in figure 8.15. Since the input sequence is persistantly exciting, the input block Hankel matrix is of full row rank $2mi = 60$. The singular values from $(2mi + 1) = 61$ on are depicted in the right hand side of figure 8.15. They allow to determine an appropriate approximate order, which for this example was chosen equal to 4. The identification was performed using the first 300 input-output data. The results of the simulations of the 6 outputs based on these data can be found in figure 8.16. Also in figure 8.16, one finds a prediction of the behavior of the system for output 1 and output 4, from time step 800 to 1000, with the aid of the identified model for the first 300 data. The prediction of output 1 was the worst while that of output 4 was the best of all predictions. Obviously, this model is very good, despite an off-set in the first output, the cause of which could be further investigated.

8.7 Conclusions.

In this chapter, we have discussed how linear dynamic models can be obtained from measurements on a multivariable system. In a first part, we have developed some new input-output equations in terms of input-output block Hankel matrices. Besides some persistancy of excitation results, we have analysed in depth the rank property of these input-output block Hankel matrices. Attention was paid to the mathematical characterization of causal dependency between variables.
Based upon the derived insights, we have discussed several identification techniques. Some of them were heuristic (least squares and total least squares analogons), others can be shown

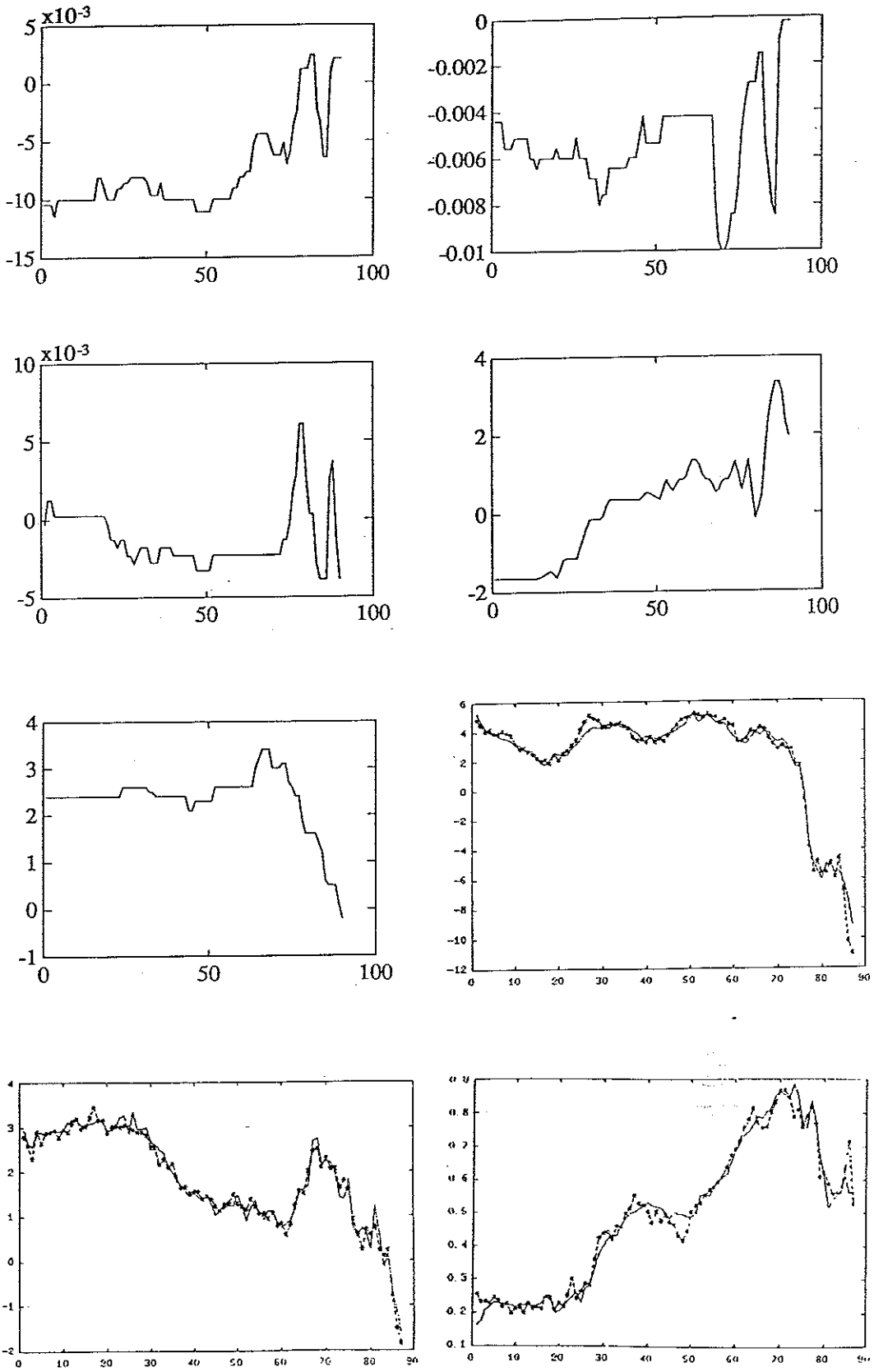


Figure 8.14: Identification of Ethane-Ethylene Distillation Column, 5 inputs, 3 outputs. (Measured: full; Simulation: Stars.)

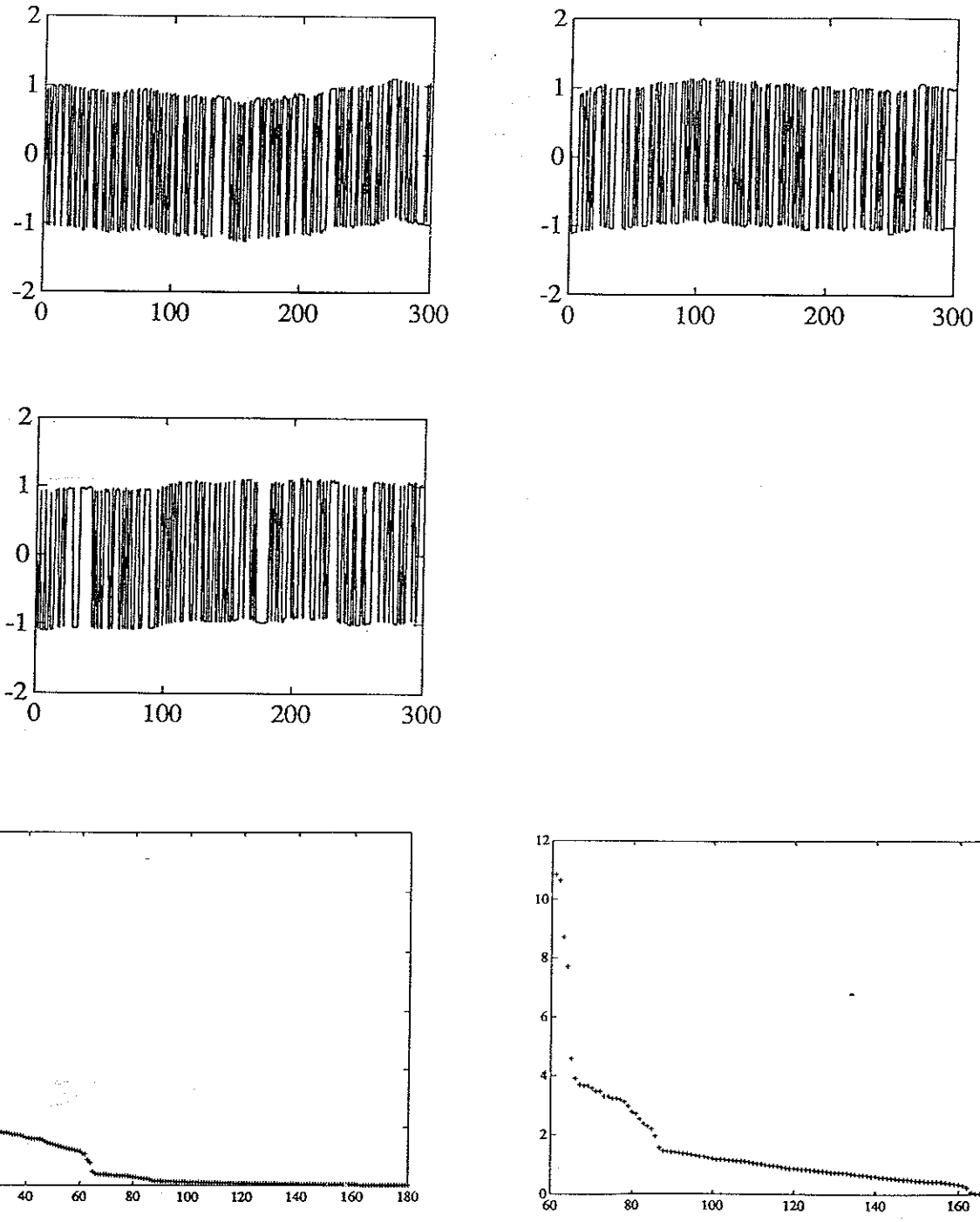


Figure 8.15: 3 inputs of feeder, singular spectra of input-output block Hankel matrix.

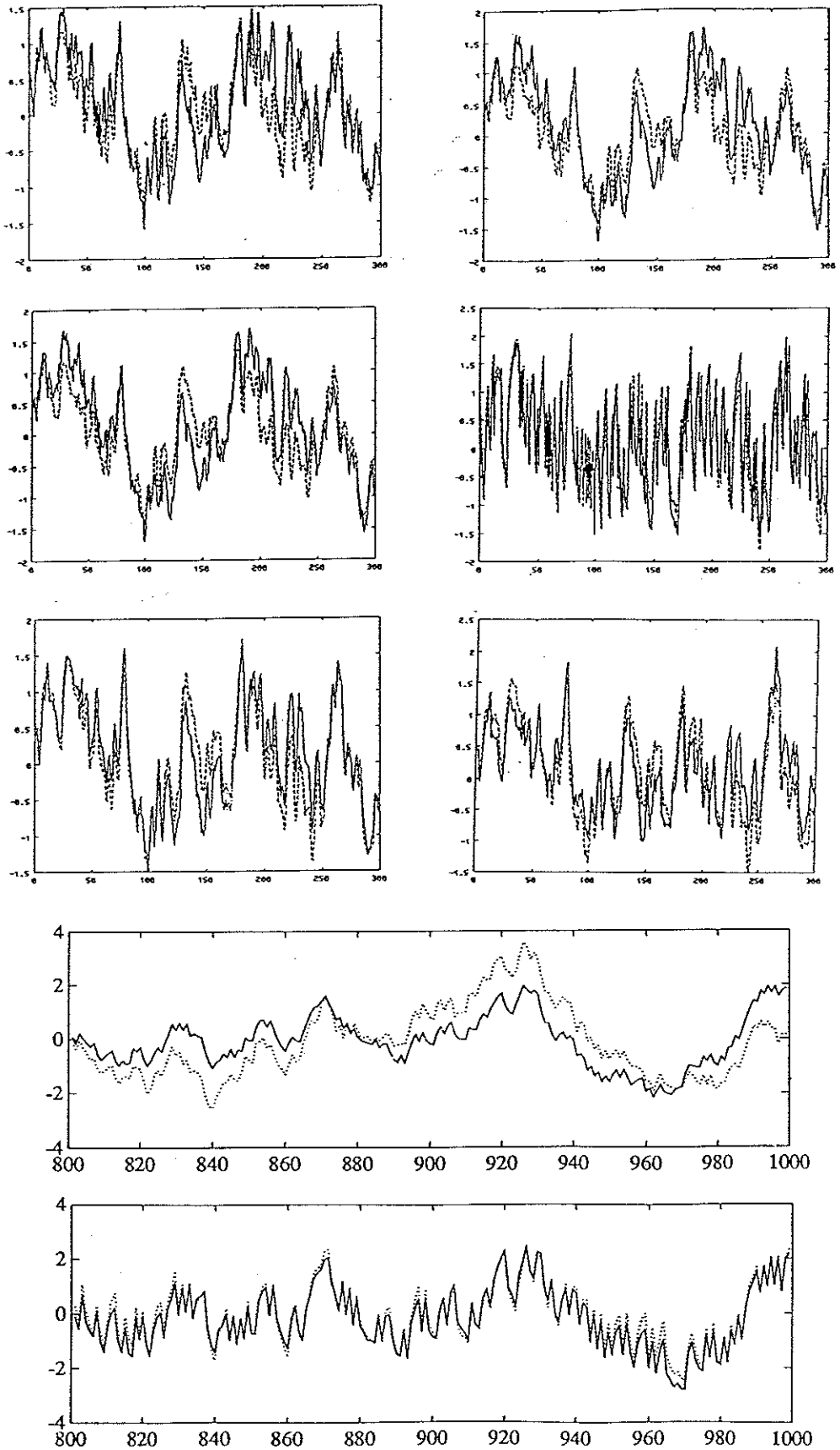


Figure 8.16: Simulation of 6 outputs (measured:full line) and prediction of output 1 and 4 (measured:full line)

to minimize a certain criterion with as a result an approximate state sequence realization. In a second step one then solves a set of linear equations in order to obtain the state space matrices and the residuals. The (generalized) singular value decomposition is a key tool in these approaches together with the notion of angles between subspaces. It was shown how a reduction to the so-called short space, allows to reduce the computational complexity.
Finally, the results were illustrated with the identification, simulation and prediction of measurements from several industrial plants.

Bibliography

[1] Adamjan V.M., Arov D.Z., Krein M.G. *Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur/Takaji problem.* Transl. Math. USSR Sbornik, vol.15, no.10,pp.31-73, 1978.

[2] Akaike A. *Markovian Representation of Stochastic Processes by Canonical Variables.* SIAM J. Control 13, pp.162-173, 1975.

[3] Akaike A. *Stochastic Theory of Minimal Realization.* IEEE Trans. on Automatic Control., Vol.19, pp.667-674, 1974.

[4] Akaike A. *A New Look at the Statistical Model Identification.* IEEE Trans. on Automatic Control, Vol.AC-19, no.6, December 1974, pp.716-722.

[5] Arun K.S., Kung S.Y. *Generalized Principal Components Analysis and its Application in Approximate Stochastic Realization.* in Modelling and Application of Stochastic Processes, Edited by U.B. Desai, Kluwer Academic Publishers, 1986, pp.75- 103.

[6] Astrom K., Eykhoff P. *System Identification: A survey.* Automatica, Vol.7., p.123-162,

[7] Astrom K.J., Wittenmarck B. *Computer Controlled Systems, Theory and Design.* Prentice Hall Int., Inc. 1984.

[8] Backx T. *Identification of an Industrial Process: A Markov parameter approach.* Doct. Diss., Techn. Univ. Eindhoven, November 1987.

[9] Baram Y. *Realization and Reduction of Markovian models from nonstationary data.* IEEE Trans. Automatic Control, Vol.AC-26, no.6, pp.1225-1231, 1981.

[10] Budin. *Minimal Realization of Discrete Linear Systems from input-output observations.* IEEE Trans. on Automatic Control, Vol AC-16, no.5, October 1971.

[11] Davis F.J., Hersh R. *L'univers Mathematiques.* Gauthier-Villars, 1985.

[12] De Moor B., Moonen M., Vandenberghe L., Vandewalle J. *Identification of Linear State Space Models with Singular Value Decompoisition Using Canonical Correlation Analysis.* Proc. of the International Workshop on SVD and Signal Processing. Grenoble, France. Ed. North Holland, Elseviers Science Publishers (in press).

[13] De Moor B., Moonen M., Vandenberghe L., Vandewalle J. *A geometrical approach for the identification of state space models with singular value decomposition.* Proc. ICASSP, New York, April 1988

[14] Eykhoff P. *System Identification.* Academic Press, 1974.

[15] Gantmacher F.R. *Théorie des Matrices.* Tome 1/2, Dunod, Paris, 1966.

[16] Golub G., Van Loan C. *Matrix Computations.*, North Oxford Academic Publishing Co., John Hopkins University Press, 1983.

[17] Goodwin G.C., Sin K.S. *Adaptive filtering.* Prentice Hall, Information and System Sciences, T. Kailath (Series Editor), Englewood Cliffs, NJ, 1984.

[18] Gopinath. *On the Identification of Linear Time-invariant Systems from input-output data.* Bell System Technical Journal, Vol.48, no.5, june 1969.

[19] Guidorzi R. *Invariants and Canonical Forms for Systems Structural and Parametric Identification.* Automatica, Vol.17, no.1, 1981.

[20] Guidorzi R., Rossi R. *Identification of a power plant from normal operating records.* Automatic Control Theory and Applications, Vol.2, No.3, pp.63-67, 1974.

[21] Guidorzi R. *On the last eigenvalue test in the structural identification of linear multivariable systems.* Proc. V-th European meeting on Cybernetics and Systems Research, Vienna, April 1980.

[22] Kailath T. *A view of three decades of linear filter theory.* IEEE Trans. Information Theory, Vol.20, pp.146-181, 1974.

[23] Kailath T. *Linear Systems.* Prentice Hall, Information and Systems Sciences, 1980.

[24] Kalman R.E. *A new approach to linear filtering and prediction problems.* J. Basic Eng. 82, pp.35-35, March 1960.

[25] Klema V.C., Laub A.J. *The singular value decomposition: its computation and some applications.* IEEE Trans. Automatic Control, Vol.AC-25, 1980.

[26] Kung S.Y. *A new identification and model reduction algorithm via singular value decomposition.* Proc. 12th Asilomar Conf. on Circuits, Systems and Computers, Pacific Grove, pp.705-714, 1978.

[27] Larimore W.E. *System Identification, Reduced Order Filtering and Modelling via Canonical Variate Analysis.* Proc. of 1983 Automatic Control Conference, San Francisco, 1984.

[28] Lawson C., Hanson R.*Solving least squares problems.* Prentice Hall Series in Automatic Computation, Englewood Cliffs, 1974.

[29] Liu, Suen. *Minimal Dimension Realization and Identification of Input-Output Sequences.* IEEE Trans. on Automatic Control, Vol.AC. April, 1977 + Comments in Vol. AC-23, no.4, August 1978 + Correction Vol. AC-23, no.1, February 1978.

[30] Liu, Suen. *Determination of Structure of Multivariable Stochastic Linear Systems.* IEEE Trans. Automatic Control, Vol.AC-23, no.3, June 1978.

[31] Ljung L., Soderstrom T. *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, Massachusetts, 1983.

[32] Moonen M., De Moor B., Vandenberghe L., Vandewalle J. *On- and off-line identification of linear state space models.* Accepted for Int. J. Control, 1988.

[33] Moore B.C. *Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction.* IEEE Trans. on Automatic Control, Vol.AC-26, no.1, February 1981.

[34] Norton J.P. *An introduction to Identification.* Academic Press, 1986.

[35] Papoulis A. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill Series in Electrical Engineering, McGraw-Hill Inc, 1984.

[36] Ramos J.A., Verriest E.I. *A unifying tool for comparing stochastic realization algorithms and model reduction techniques.* Proc. of the 1984 Automatic Control Conference, San Diego, CA, June 1984.

[37] Rissanen J.*Predictive and non-predictive minimum description length principles.* Time series and Linear Systems. Lecture Notes in Control and Information Systems, Springer Verlag, 1986. (ed. Sergio Bittanti).

[38] Staar J. *Concepts for reliable modelling of Linear Systems with apllication to on-line identification of multivariable state space descriptions.* Doct. Diss., Departement Elektrotechniek, Katholieke Universiteit Leuven, June 1982.

[39] Swevers J., Adams M., De Moor B. *A new direct time domain identification method for linear systems using singular value decomposition.* Proc.12-th Int. Sem. Modal Analysis, K.U. Leuven, 21-24 september 1987.

[40] Tan S. *A unified approach for analysis and control of multivariable non-causal systems.* Doct. Diss., Departement Elektrotechniek Katholieke Universeit Leuven, October 1987.

[41] Tether A.J.*Constructing of minimal linear state variable models from finite input-output data.* IEEE Trans. Automatic Control, Vol.AC-15, no.4, pp.427-436 (1970).

[42] Vandenberghe L, Van Miegem P. *Een nieuw algoritme voor de identifikatie van lineaire multivariabele systemen.* Master Thesis, ESAT, Departement Elektrotechniek, Katholieke Universiteit Leuven, 1987.

[43] VanHuffel S.,*Analysis of the total least squares problem and its use in parameter estimation.* Doctoral Thesis, Katholieke Universiteit Leuven, Department Elektrotechniek, Juni 1987.

[44] Verhaegen M. *A new class of algorithms in linear system theory.* Doct.Diss., Departement Elektrotechniek, Katholieke Universiteit Leuven, November 1985.

[45] Verriest E.I., Kailath T. *A generalized balanced realization.* IEEE Trans. Automatic Control, Vol.AC-28, no.8, pp.833-844, 1983.

[46] Verriest E.I. *A unified theory of model reduction via Gleason measures.* IMA Conference on Mathematics in Signal Processing, University of Bath, September 1985, United Kingdom.

[47] Wertz V. *Structure selection for the identification of multivariate processes.* Doct. Diss., Université Catholique de Louvain, May 1982.

[48] Willems J.C. *From time series to linear systems.* Automatica, Part I:vol.22, no.5, pp.561-580, 1986. Part II:vol.22, no.6, pp.675-694, 1986. Part III:vol.23, no.1, pp.87-115, 1987.

[49] Wonham M. *Linear Multivariable Control, A Geometric Approach.* New York Springer, 1974.

Chapter 9

Conclusions and Perspectives

In this doctoral dissertation, we have developed and discussed *concepts and algorithms for modelling of static and dynamic systems.*

The presented results can be divided in 4 parts, in which we shall emphasize our own contributions.

1. Nonnegative linear algebra and the generalized linear complementarity problem (chapter 2 and 3).
2. Oriented energy and oriented signal-to-signal ratio concepts and algorithms (chapter 4).
3. Identification of linear relations for static systems (chapter 5) and the treatment of an uncertainty principle of mathematical modelling for the case of linear relations between variables (chapter 6).
4. Concepts and algorithms for the analysis and identification of dynamic linear systems (chapter 7 and 8).

Of course, this systematic summary of results contains *implicitly* numerous suggestions for further research. Instead of repeating them explicitly, it is tacitly assumed that the almost *unlimited* number of future possibilities, follows immediately from the current status, as reported in this dissertation and its conclusions.

Nonnegative linear algebra and the generalized linear complementarity problem.

Starting from the 'classical' linear complementarity problem, we have formulated in chapter 3, a generalized version, which we have baptized the *generalized linear complementarity problem* and we have derived an *algorithm* to find all of its solutions. The epithet *generalized* can be understood from the following considerations:

- The classical linear complementarity problem reduces to a special case of the presented generalized version.
- The complementarity conditions may be much more complicated than a simple combination of nonnegativity and orthogonality. As a matter of fact, the most general form

of the complementarity conditions consists of a sum of products of components of the solution vector(s).

- The class of systems and behaviors that may be described in the new framework is much larger than that of the classical problem formulation.

The algorithm, derived in chapter 3, is based upon a threefold induction. Two of them have been studied in detail in chapter 2 in the context of finding all nonnegative solutions to a set of linear (in-)equalities. The third one arises from the complementarity conditions.

1. The first induction allows to find all nonnegative vectors that are orthogonal to a set of given vectors. The algorithm essentially consists of updating the old solution set for every new given vector.

2. The second induction allows for a redundancy reduction, in each update, such that only the extremal vertices of the polyhedral solution set are obtained. Hereto, necessary and sufficient conditions for extremity and adjacency have been derived.

3. The third induction consists of an elimination in each update, of those extremal vertices of the polyhedral solution cone, that do not satisfy the complementarity conditions.

The main features of the algorithm are the following:

- The proof of the validity of this algorithm at the same time demonstrates that the solution set of the generalized linear complementarity problem consists of the union of (un-)bounded polyhedra (discrete vectors, polyhedra and polyhedral cones).

- The algorithm finds *all* solutions, contrary to algorithms for the 'classical' linear complementarity problem.

- Because of the fact that no matrix inversions or other possibly ill-conditioned matrix operations are needed, the algorithm allows for a numerical reliable implementation.

- By a verification of the so-called *cross-complementarity conditions*, one can check which nonnegative combinations of solutions are solutions as well. This allows for a complete geometrical characterization of the solution set.

A lot of applications were discussed:

- It was shown how the GLCP can solve a lot of geometrical problems such as problems with piecewise linear descriptions, in 2 and more dimensions and the intersection of geometrical objects. Several parametrization strategies have been proposed.

- The inimate relation with mathematical programming has been established, especially quadratic optimization.

- The analysis of piecewise linear resistive networks is most easily performed in terms of the GLCP: It was shown how our approach allows to determine *all* solutions where the classical approaches fail to do this.

- The computation of *all* invariant states of neural networks, with and without a priori partial information, is a GLCP problem.

The computational complexity may be rather involved, which can be explained by the fact that the solution set itself may consist of a considerable amount of extreme vertices and the fact that the number of intermediate vertices heavily depends upon the order in which the equations are treated.

Oriented energy and oriented signal-to-signal ratios.

In a lot of applications, two or more vector sequences of observations are to be compared. Typically, these measurements are arranged in overdetermined matrices with, say, many more columns than rows. The oriented energy of one vector sequence is nothing else but the spatial distribution of the energy in all directions of the ambient space. There is a direction of strongest and one of weakest energy and there are intermediate 'saddle-points'. The singular value decomposition allows to quantify both the extremal directions and the extremal oriented energies. When the relative oriented energy distribution of two vector sequences is to be compared, it is demonstrated how the generalized singulare value decomposition allows to quantify both geometrically and numerically the oriented signal-to-signal ratio distribution. In each direction of the ambient space, the oriented energy of the first vector sequence is compared with that of the other one.
We have provided several illustrations of this conceptual framework, including Gauss Markov (prewhitening) estimators, total linear least squares, realization theory and the identification of linear dynamical systems and several applications of factor-analysis such as high resolution signal separation and the separation of maternal from fetal ECG.

Identification of linear relations in noisy data

Deductive versus inspirational modelling approaches.

There exists a close connection between the notions of *orthogonality* and *linear combinations* and the distinction between what is considered to be a *model* and what is considered to be *noise.*
This leads to the distinction between 2 possible modelling philospohies: The deductive versus the inspirational approach. *Deductive modelling* essentially reduces to a *perturbation analysis.* It is assumed that the exact data are corrupted by unknown perturbations, that are characterized via some global measures such as e.g. probability density functions. It is then investigated how these perturbations could have affected the exact data.
The *inspirational point of view* essentially reduces to *approximation*: How well can the observations be approximated by a model from a certain model class, optimizing a certain criterion? The question of *consistency* for the deductive approach concentrates around the study of the conditions under which the original exact data can be recovered or modelled. In the inpiration approach, *consistency* is that property of an identification scheme that states than, when the data are exact, the exact model is found.

Deductive identification of linear relations.

Two deductive key theorems have been derived and proved: The *orthogonality theorem* and the *lever theorem.*

In the orthogonality theorem, we have derived the probability density functions of the distributions of the angle between an 'exact' vector and a vector, of which the compononents are *spherically* distributed. The result is an explicit demonstration of the fact that 'exact' data and 'pure noise' data tend to be more and more orthogonal (with respect to the Euclidean inner product), as the number of observations is increased.
The main result contained in the lever theorem is the demonstration of the *consistency* of the *short* space of a matrix, when perturbed by additive noise. It is shown how the *the long space is irreversibly* lost and an asymptotic expression is derived.

These two theorems allow for considerable insight in *classical identification schemes* and the *biasedness* of the estimation of *canonical angles* between subspaces. Moreover, they explain the necessity of the so-called *Mahalanobis* transformation, which is equivalent with the use of an inner product, with as weighting matrix the inverse of the noise covariance matrix. This relates the lever theorem to the oriented energy framework of chapter 4.

It is shown how classical identification techniques, such as *linear least squares* and *total linear least squares* are special cases of one theorem, which analysis the rank reduction of matrices via *rank one modifications.* It is shown how *all* least squares solutions of an identification problem, are the column vectors of the inverse of the Grammian of the data matrix. The total least squares solution, lies within the polyhedral cone of these least squares solutions, at least for sufficiently high signal-to-noise ratios. Furthermore, it is demonstrated that the noise model that is delivered by total least squares, is completely unsatisfactory. Fortunately, the total least squares solutions coincides with that of the *identity matrix approach*, which delivers a noise model that is more 'realistic'.

The notion of *canonical angles between subspaces* is a generalization of the angle between 2 vectors. It is shown how the generalized singular value decomposition allows the computation of an approximate intersection between subspaces, in case the data are perturbed by noise. Hereto several criteria are proposed. Some of them ultimately result in a *canonical correlation analysis*, while others are more explicitly based upon the *generalized singular value decomposition.* The lever theorem allows to compute the *asymptotic bias* on the subspaces as a function of the noise variance. This result is confirmed by simulations. The problem with canonical correlation is the explicit normalization, which treats strong and weak components of the signals on the same level. Thereto, a new criterion is analysed that takes explicitly into account, the relative energy power of the signal modes. Its computation alse uses the generalized singular value decomposition.

The uncertainty principle of mathematical modelling.

In chapter 6, we have discussed what is called *an uncertainty principle of mathematical modelling.* In essence it reduces to the observation that, when data are not compatible with a certain model (due to several possible causes such as noise, non-linearities, etc...), any model that aims at explaining the data, must be characterized by some uncertainty that reflects the incompatibility of the model with the data. At the same time, the principle states that identification techniques that deliver *unique* models for noisy data, sooner or later have to invoke assumptions, that are a priori unverifiable, in order to obtain the required *uniqueness* of the result.

In chapter 6, it is shown how the problem of identifying linear relations from noisy data, can be formulated mathematically, as the minimization of the rank of a symmetric positive matrix, when only its diagonal elements may be changed, subject to nonnegative definiteness conditions for the result. This is the so-called *Frisch scheme.* In an historical review, it is shown how all existing results concentrate around the *algebraic* analysis of the problem, such as determinantal relations, algebraic nonnegative definiteness conditions via principal minors or by employing the celebrated Perron-Frobeniustheorem.
Our approach concentrates upon the *geometrical* nature of the problem. It is shown how candidate solution vectors have to satisfy three necessary conditions: They must be *orthant invariant*, *null invariant* and *allowed.* Several additional results allow to determine the minimal rank. For instance, it is shown that the minimal rank is r, if there exist at least $n - r$ (n is the number of variables), linearly independent allowed orthant null invariant vectors, each with at least $n - r - 1$ zeros. It is shown how the algorithm of chapter 2, which finds all nonnegative solutions to a set of linear equations, can be employed to find all allowed orthant null invariant vectors with zeros. Hence, *the maximal corank can be determined.*
We also briefly discuss the *generic solvability* of the problem, which results in the Wilson-Ledermann bound, which is a lower bound on the minimal rank that is generically achievable.

The finishing touch to the complete solution is however still lacking: More specifically, it is conjectured that the solution set consists of the union of polyhedral cones of allowed orthant null invariant vectors, the vertices of which are completely determined by our algorithm. These polyhedral cones characterize the incompatibility of the raw data with the assumed static linearity of the model.

Identification of linear state space models for dynamic systems.

The main results concerning dynamic systems, are the analysis in depth of the so-called *shift structure* of certain subspaces (chapter 7), and based upon this shift structure, the identification of state space models from input-output measurements (chapter 8.)

Shift structure and dynamic systems.

In chapter 7, the close connection between the *shift structure* of certain subspaces and linear time invariant discrete time systems is established. It is shown how the singular value decomposition of the block Hankel matrix with Markov parameters of a system allows to obtain realization algorithms, quantitative measures for controllability and observability and a rationale for model reduction. It is shown how the shift structure of both column and row spaces of rank deficient block Hankel matrices, reveals itself in any factorization of the block Hankel matrix, including the singular value decomposition. Using non-conventional matrix calculus (Kronecker and Khatri-Rao product), it is shown how one can derive a parametrization of all *rank deficient block Hankel matrices of finite dimensions*, that share (a subset of) a certain set of minimal system poles. All of these matrices, can be found from the kernel of a certain matrix.
It is shown how these results can be used to study properties of orthonormal block Hankel

matrices and how they can be generalized to matrices without explicit structure but with an implicit shift structure. It is shown how one may recover, in a certain unstructured subspace, the subspace of maximal dimension with shift structure. This results suggests factor analysis applications that exploit the shift structure instead of the relative signal-to-signal ratio.

Identification of state space models.

In chapter 8, we have first derived some important, though less known properties of 'exact' state space models. Starting from a fundamental structured input-output equation, it was shown how the rank of the concatenation of input-output block Hankel matrices can be written as a sum of three terms, that were analysed in depth. Moroever, we have paid attention to the mathematical characterization of causal dependency and the influence of delays on the singular values of the input-output block Hankel matrix.

Based upon these insights, we have derived several identification strategies. All of them contain a geometrical idea and are translated in numerical reliable algorithms, essentially based upon the (generalized) singular value decomposition. We have discussed a *heuristic* linear and total linear least squares identification approach, together with a deductive analysis of the perturbing influence of additive noise. Other identification methods follow from the observation that a state sequence can be realized from an intersection of the row spaces of past and future input block Hankel matrices. Again, several versions are possible, such as linear and total linear least squares and forward and backward prediction approaches, all of which are essentially based upon canonical angles between subspaces. We also derive an unnormalized predictive solution strategy. The system matrices are computed via the solution of a set of linear equations and the reduction to the so-called *short* space allows for a considerable reduction in computational complexity, leading to an on-line implementation. Finally, we have illustrated the theory with several practical applications of the identification of industrial processes.

Besides further research for optimization of the computational complexity via adaptive (G)SVD algorithms, an important *conceptual* open problem concerns the optimality of the computed state sequence and its relation with the state sequence as estimated by the Kalman filter, once a state space model is known.

Appendix B: The Kronecker and Khatri-Rao product.

There are numerous applications of the Kronecker product in various fields including statistics, economics, optimisation and control. One of the major advantages of the matrix calculus of Kronecker and Khatri-Rao products is that it simplifies considerably complicated calculations e.g. derivates of a matrix with respect to another matrix. Interesting introductions can be found among the references of chapter 7.

Definition 1 Kronecker product.
The Kronecker product of a matrix A (p timesq) and a matrix B $(m \times n)$, denoted by $A \otimes B$, is the $(pm) \times (qn)$ matrix defined as:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1q}B \\ a_{21}B & a_{22}B & \dots & a_{2q}B \\ \dots & \dots & \dots & \dots \\ a_{p1}B & a_{p2}B & \dots & a_{pq}B \end{pmatrix}$$

Definition 2 Khatri-Rao product.
The Kathri-Rao product of two matrices G $(s \times u)$ and F $(t \times u)$ (with the same number of columns !), denoted by $F \odot G$, is a $(st) \times u$ matrix defined by:

$$F \odot G = [f^1 \otimes g^1 \; f^2 \otimes g^2 \; \dots \; f^u \otimes g^u]$$

where f^i and g^i, $i = 1, \dots, u$ are the columns of F and G.

The Kathri-Rao product is nothing more than the Kronecker product columnwise.
Example:

$$F = \begin{pmatrix} 3 & 1 \\ 2 & -1 \\ 0 & 2 \end{pmatrix} \quad G = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$F \odot G = \begin{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \\ 2 & 0 \\ 0 & -2 \\ 0 & 0 \\ 0 & 4 \end{pmatrix}$$

Let A be a $m \times n$ matrix with columns a^i. Then an important vector valued function is

$$vec(A) = \begin{pmatrix} a^1 \\ a^2 \\ \cdots \\ a^n \end{pmatrix}$$

obtained by storing the columns a^i of A in a long column vector. For A being square $m \times m$, let $vecd(A)$ be a $m \times 1$ vector containing the diagonal elements of A. In what follows, A, B, C, D are real matrices of appropriate dimensions.

Lemma 1 Mixed product rule for the Kronecker product.

$$[A \otimes B][C \otimes D] = [AC] \otimes [BD]$$

Lemma 2 Vector function of a product.

$$vec[ADB] = [B^t \otimes A]vec(D)$$

Lemma 3 Mixed product rule for the Khatri-Rao product.

$$[A \otimes B][C \odot D] = [AC] \odot [BD]$$

Lemma 4 Vectorfunction of a product with a diagonal matrix.

$$vec[ADB] = [B^t \odot A]vecd(D)$$

if D is square diagonal.

Corollary 1 Vectorfunction of the singular value decomposition.
Let the SVD of A be $A = USV^t$. Then

$$vec(A) = [V \odot U]vecd(S)$$

Lemma 5 Singular value decomposition of the Kronecker product.
If $A = U_a S_a V_a^t$ and $B = U_b S_b V_b^t$ are the SVD of A and B, then the SVD of $A \otimes B$ is given by:

$$A \otimes B = (U_a \otimes U_b)(S_a \otimes S_b)(V_a \otimes V_b)^t$$

Corollary 2 *If $rank(A) = r_A$ and $rank(B) = r_B$, then $rank([A \otimes B]) = r_A r_B$*

Appendix C : Matrix lemmas

If A^{-1} occurs in an expression, it is tacitly assumed that A is square and nonsingular. It is assumed as well that matrices are compatible for the addition and multiplication where this is required.

Matrix Inversion Lemmas

Lemma 1 *Inverse of a partioned matrix*
If:

$$M = \begin{pmatrix} A & D \\ C & B \end{pmatrix}$$

then:

$$M^{-1} = \begin{pmatrix} (A - DB^{-1}C)^{-1} & -A^{-1}D(B - CA^{-1}D)^{-1} \\ -(B - CA^{-1}D)^{-1}CA^{-1} & (B - CA^{-1}D)^{-1} \end{pmatrix}$$

The matrix $B - CA^{-1}D$ is termed the *Schur complement* of the matrix B in the matrix M.

Lemma 2

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1}$$

Corollary 1 Inverse of the sum of two matrices

$$[A + B]^{-1} = A^{-1} - A^{-1}B[I + A^{-1}B]^{-1}A^{-1}$$

Corollary 2 The Woodbury formula

$$(A + UV^t)^{-1} = A^{-1} - A^{-1}U(I + V^tA^{-1}U)^{-1}V^tA^{-1}$$

Corollary 3 The Sherman-Morrison formula.
Let A be square non-singular and u and v 2 vectors such that $v^tA^{-1}u \neq -1$, then:

$$[A + uv^t] = A^{-1} - \frac{A^{-1}uv^tA^{-1}}{1 + v^tA^{-1}u}$$

Determinantal relations

Lemma 3 Schur's determinantal formula
Consider a square partioned matrix

$$M = \begin{pmatrix} A & D \\ C & B \end{pmatrix}$$

If A is square nonsingular, then :

$$det(M) = det(A)det(B - CA^{-1}D)$$

Corollary 4 *Consider a symmetrix matrix M partitioned as*

$$M = \begin{pmatrix} A & b \\ b^t & c \end{pmatrix}$$

where b and c are vectors of appropriate dimensions. If A is square nonsingular, then:

$$det(M) = det(A)det(c - b^t A^{-1}b)$$

Eigenvalue decomposition theorems.

A matrix A is *positive definite* (**PD**) if the quadratic form $x^t Ax > 0$ for all x. It is *nonnegative definite* (**NND**) if the quadratic form $x^t Ax \geq 0$ for all x. A symmetric matrix A is PD (NND) if and only if all of its eigenvalues are positive (nonnegative). A symmetric matrix A is PD if and only if the leading principal minors are all positive. On the other hand, one should check *all* principal minors to be nonnegative, in order to conclude that the matrix is NND.
A matrix A is called *positive* if all of its elements are positive.

Theorem 1 The Perron-Frobeniustheorem
If A is positive, then there exists a $\lambda_0 > 0$ and a vector $x_0 > 0$ such that:

1. $Ax_0 = x_0\lambda_0$
2. *If $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $\mid \lambda \mid < \lambda_0$*
3. *λ_0 is an eigenvalue of geometric and algebraic multiplicity 1.*

This remarkable and important theorem may be extended to nonnegative matrices. The conclusion is however weaker. A matrix is *nonnegative* if all of its elements are nonnegative.

Theorem 2 *Let A be nonnegative. Then there exists $\lambda_0 \geq 0$ and $x_0 \geq 0$ such that:*

1. $Ax_0 = x_0\lambda_0$.
2. *if $\lambda \neq \lambda_0$ is any other eigenvalue of A, then $\mid \lambda \mid \leq \lambda_0$.*

Appendix D: Proof of the Orthogonality Theorem

In this appendix [1], we compute the probability that a random vector v in $\mathcal{R}^j$ makes an angle between α and $\alpha + d\alpha$ with a fixed i-dimensional subspace $\mathcal{S}^i$. The main result is the following:

Theorem 1 *If*

- $p(\alpha, i, j)d\alpha$ *is the probability that an arbitrary direction in* $\mathcal{R}^j$ *makes an angle between* α *and* $\alpha + d\alpha$ *with a fixed* i*-dimensional subspace* $\mathcal{S}^i$.
- $v(\alpha, i, j,)d\alpha$ *is the infinitesimal small volume in* $\mathcal{R}^j$*, generated by all vectors with a norm equal or smaller than one, that have an angle between* α *and* $\alpha + d\alpha$ *with* $\mathcal{S}^i$
- V_j *is the volume of a hypersphere of radius* R *in* j *dimensions.*

then, the required probability distribution is given by:

$$p(\alpha, i, j) = v(\alpha, i, j)/V_j$$

The proof of the theorem will now proceed by a careful calculation of the required volumes V_j and $v(\alpha, i, j,)$.

The volume of a hypersphere

In the Euclidean vectorspace of all real j-tuples $\mathcal{R}^j$, the j-dimensional unit hypersphere is defined as $\mathcal{S}^j = \{x \in \mathcal{R}^j \mid x^t x = 1\}$. The surface of such a hypersphere has $j-1$ dimensions. In general the axis of rotation symmetry is a $(j-2)$-dimensional subspace (but be careful in generalization of 2 and 3 dimensions: a 4-dimensional hypersphere can rotate simultaneously around 2 fixed planes that are orthogonal). If an j-dimensional hypersphere is projected on a $(j-1)$-dimensional subspace, the result is a $(j-1)$-dimensional hypersphere. The same is true for the intersection of an j-dimensional hypersphere with an $(j-1)$-dimensional subspace: The result is a $(j-1)$-dimensional hypersphere. For intersections with lower dimensional subspace, the same attention is necessary as with rotational symmetry: A 4-dimensional sphere cannot be cut by a 2-dimensional plane in 2 separate pieces: it remains one piece! If a hypersphere is made of sufficiently flexible material, it can be turned inside out by passing via a space with one more dimension. But one of the most unexpected and extraordinary results (that will be explicitly derived here), is the fact that the volume V^j of the j-dimensional

[1] I am indebted to Piet Van Mieghem en Lieven Vandenberghe for the technical derivation of the result.

hypersphere, decreases as a function of j for $j \geq 5$. If $j \to \infty$, then $V^j \to 0$. It is possible to place any object, how big and large it might be, in a j-dimensional hypercube where j is sufficiently large: This can easily be seen from the fact that the distance between 2 opposite cornerpoints of a j-dimensional cube is $\sqrt{j}$ and there is no supremum for the row $\sqrt{j}, j \to \infty$. But a j-dimensional sphere can never contain an object which has a length larger than 2 times its radius, how big j might be.

Theorem 2 *The volume of a j-dimensional hypersphere with radius R equals:*
For even j:

$$V_j = \frac{\pi^{j/2}}{(1/2j)!} R^j$$

For odd j:

$$V_j = \frac{2^j \pi^{(j-1)/2}((j-1)/2)!}{j!} R^j$$

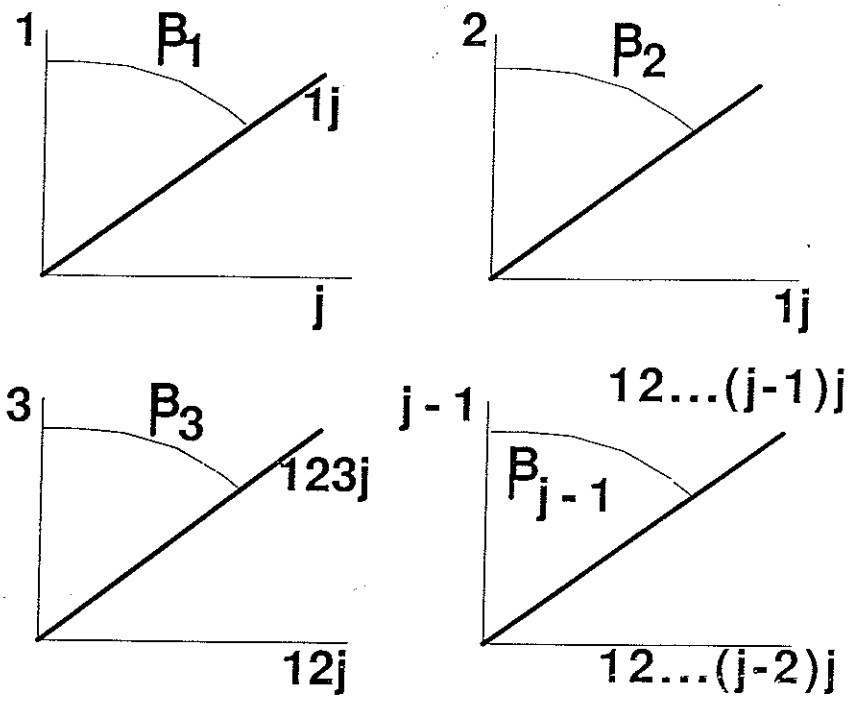
Proof: The elegance of the proof depends upon the appropriate choice of a set of coordinates. Any vector v within $\mathcal{R}^j$ will be parametrized by its length r and by $j-1$ angles $\beta_1, \ldots, \beta_{j-1}$, to be determined as follows:

- Choose an orthonormal set of j basis vectors e_1 , ... , e_j.
- Project the vector v onto $span(e_1, e_j)$ and call the projection v_{1j}. Call β_1 the angle between e_1 and v_{1j}.
- Project the vector v onto $span(e_2, v_{1j})$ and call the projection v_{12j}. The angle between e_2 and v_{12j} is called β_2.
- Project the vector v onto $span(e_3, v_{12j})$ and call the projection v_{123j}. This defines β_3 as the angle between e_3 and v_{123j}.
- .....
- Project the vector v onto $span(e_{j-1}, v_{12\ldots(j-2)j})$. Since the vector v lies in this plane, it equals its own projection: Hence $v_{123\ldots(j-2)(j-1)j} = v$. The angle between v and e_j is called β_{j-1}.

This procedure is depicted schematically in figure 1. This choice of coordinates allows to describe an infinitesimal small volume around v, containing all vectors in the 'box' formed by $dr, d\beta_1, d\beta_2, \ldots, d\beta_{j-1}$. This will permit to compute the volume of the hypersphere. One has simply got to go the way back, starting from the plane $span(e_{j-1}, v_{12\ldots(j-2)j})$ to the plane $span(e_1, e_j)$. This infinitesimal volume around v is described by:

$$\begin{aligned} dV_j &= [(r\beta_{j-1})dr][rsin(\beta_{j-1})d\beta_{j-2}][rsin(\beta_{j-1})sin(\beta_{j-2})d\beta_{j-3}] \\ &\quad .....[rsin(\beta_{j-1})sin(\beta_{j-2})\ldots sin(\beta_2)d\beta_1] \\ &= r^{j-1}(sin(\beta_{j-1}))^{j-2}(sin(\beta_{j-2}))^{j-3}.....(sin(\beta_2))dr\, d\beta_{j-1}\ldots d\beta_1 \end{aligned}$$

if $j \geq 2$. Observe that $dV_1 = dr$ and that for $j = 3$ we obtain the classical three dimensional polar coordinates. The volume of the hypersphere can now be computed as the multiple integral of this expresion, whereby:



Choice of coordinates

- r varies between 0 and R.
- β_{j-1} varies between 0 and 2π.
- β_{j-2} varies between 0 and π.
- ...
- β_1 varies between 0 and π.

Then:

$$\begin{aligned} V_j &= 2\int_{r=0}^{R}\int_{\beta_1}^{\pi}\int_{\beta_2}^{\pi}\cdots\int_{\beta_{j-2}}^{\pi} r^{j-1}(sin\beta_{j-1})^{j-2}(sin\beta_{j-2})^{j-3}\ldots sin\beta_2 dr\, d\beta_{j-1}\ldots d\beta_1 \\ &= 2(R^j/j)A_{j-2}A_{j-3}\ldots A_1A_0 \\ &= 2(R^j/j)C_j \end{aligned}$$

for $j \geq 2$. Here $C_j = A_0A_1\ldots A_{j-2}$ and:

$$A_i = \int_0^{\pi}(sinx)^i dx$$

The computation of these integrals can be found in standard books on integrals. First, introduce the notation $i!! = i(i-2)(i-4)\ldots 1$ for i odd and $i!! = i(i-2)(i-4)\ldots 2$ if i is

even.

$$\begin{aligned} A_{2i} &= \pi \frac{(2i-1)!!}{(2i)!!} \qquad A_0 = \pi \\ A_{2i+1} &= 2 \frac{(2i)!!}{(2i+1)!!} \qquad A_1 = 2 \end{aligned}$$

Hence, it follows that:

$$\begin{aligned} C_{2j+1} &= 2^j \frac{\pi^j}{(2j-2)!!} \qquad j \geq 1 \\ &= 1 \qquad\qquad\qquad\quad j = 0 \\ C_{2j} &= 2^{j-1} \frac{\pi^j}{(2j-2)!!} \qquad j \geq 2 \end{aligned}$$

This proves the theorem. □

Observe that the constants C_j and hence the volumes V_j can be obtained recursively:

- $C_1 = 1$
- $C_2 = \pi$
- $C_j = \frac{2\pi}{(j-2)} C_{j-2}$ for $j \geq 2$.

The surface area of a hypersphere is obtained from V_n by differentiating with respect to r. A counter-intuitive property is that, the function V_n as a function of n for $r = 1$, reaches a maximum at $n = 5$.

The volume $v(\alpha, i, j)$.

Consider a vector v inside the hypershere with radius R in j dimensions. Consider also an i dimensional subspace $\mathcal{S}^i$ and define the volume $v(i, j, \alpha)d\alpha$ as the volume inside the hypersphere, generated by all vectors that make an angle between α and $\alpha + d\alpha$ with $\mathcal{S}^i$.

Theorem 3 *The volume $v(\alpha, i, j)$ is given by:*

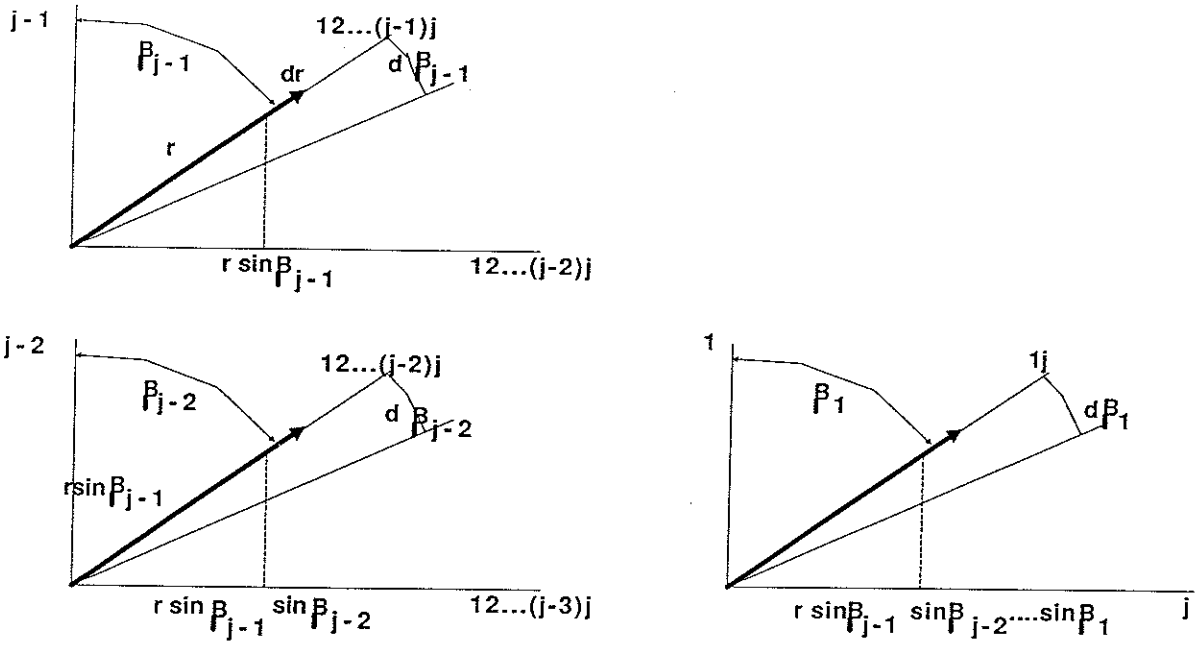
$$v(\alpha, i, j)d\alpha = \frac{4}{j} C_i C_{j-1} (cos\alpha)^{i-1} (sin\alpha)^{j-i-1} d\alpha$$

where $C_k = A_0 A_1 A_2 \ldots A_{k-2}$, $C_1 = 1$ and

$$A_k = \int_0^\pi (sinx)^k dx$$

Proof: The proof goes along the same lines as the proof of theorem 2. The vector v is described by:

1. Its norm r
2. $j - 1$ angles β_1 , β_2 , $\ldots$, β_{j-1}



Infinitesimal small volume

The angles are determined in precise the same way as described above, with this specification that now the first $i-1$ of them β_1 , ... , β_{i-1} are determined by an orthonormal set of basis vectors e_1 , ..., e_i in the subspace $\mathcal{S}^i$. The other basis vectors e_{i+1} ,... , e_j are chosen in the orthogonal complement $(\mathcal{S}^i)^\perp$ of $\mathcal{S}^i$. Then the infinitesimal small volume around the vector v can be found from inspection of figure 3 as:

if $i > 1$ and $j > (i+1)$: The volume is given by:

$$\begin{aligned} dV(i,j) &= [dr(rd\theta)][rcos(\theta)d\beta_{i-1}][rcos(\theta)sin(\beta_{i-1})d\beta_{i-1}] \\ &\quad [rcos(\theta)sin(\beta_{i-1})\dots sin(\beta_2)d\beta_1][rsin(\theta)d\beta_{j-2}][rsin(\theta)sin(\beta_{j-2})d\beta_{j-1}] \\ &\quad \dots[rsin(\theta)sin(\beta_{j-2})\dots sin(\beta_{i-1})d\beta_i] \\ &= r^{j-1}[(cos\theta)^{i-1}(sin\theta)^{j-i-1}][(sin\beta_{i-1})^{i-2}(sin\beta_{i-3})^{i-3}\dots(sin\beta_2)] \\ &\quad [(sin\beta_{j-2})^{j-i-2}(sin\beta_{j-3})^{j-i-3}\dots sin\beta_{i+1}]dr\ d\theta\ d\beta_1\dots d\beta_{j-2} \end{aligned}$$

if $i = 1$ and $j > 2$:

$$dV(i,j) = r^{j-1}(sin\theta)^{j-2}(sin\beta_{j-2})^{j-3}\dots(sin\beta_{i+1})\ dr\ d\theta\ d\beta_1\dots d\beta_{j-2}$$

if $i > 1$ and $j = i+1$:

$$dV(i,j) = r^{j-1}(cos\theta)^{i-1}(sin\beta_{i-1})^{i-2}(sin\beta_{i-2})^{i-3}\dots sin\beta_2\ dr\ d\theta\ d\beta_1\dots d\beta_{i-1}$$

if $i = 1$ and $j = 2$:

$$dV(i,j) = rdrd\theta$$

For each of these cases, the integration is then performed with the following data:

- $r \in [0,1]$
- $\theta \in [\alpha, \alpha + d\alpha]$
- $\beta_1 \in [0, 2\pi], \beta_2 \in [0, \pi], \ldots, \beta_{i-1} \in [0, \pi]$
- $\beta_i \in [0.2\pi], \beta_{i+1} \in [0, \pi], \ldots, \beta_{j-2} \in [0, \pi]$

Hence, one proves the theorem by calculating the integral:

$$v(\alpha, i, j) = 4d\alpha \int_{r=0}^{1} \int_{\beta_1=0}^{\pi} \cdots \int_{\beta_{j-2}=0}^{\pi} dV(i,j)$$

□

The probability distribution function

It is now straightforward to compute the probability distribution function $p(\alpha, i, j)d\alpha$ which is the probability that an arbitrary vector in $\mathcal{R}^j$ makes an angle between α and $\alpha + d\alpha$ with a fixed i-dimensional subspace $\mathcal{S}^i$.

$$\begin{aligned} p(\alpha, i, j) &= v(\alpha, i, j)/V_j \\ &= 2\frac{C_i C_{j-i}}{C_j}(cos\alpha)^{i-1}(sin\alpha)^{j-i-1}. \end{aligned}$$

The cumulative distribution function

Definition 1 *The cumulative distribution function $P(\alpha, i, j)$ is defined as:*

$$P(\alpha, i, j) = \int_{(\pi/2)-\alpha}^{\pi/2} p(\alpha, i, j)\, d\alpha$$

Obviously, $P(\alpha, i, j)$ is the probability that the angle between an arbitrary vector in $\mathcal{R}^j$ and $\mathcal{S}^i$ lies between $\pi/2 - \alpha$ and $\pi/2$. With the results of the preceding sections, this can be written as:

$$P(\alpha, i, j) = 2\frac{C_i C_{j-i}}{C_j} I(\alpha, i, j)$$

where

$$I(\alpha, i, j) = \int_0^{\alpha} (sinx)^{i-1}(cosx)^{j-i-1} dx$$

This integral can be computed by employing the following recursions:

1. **Initialisation :** If j is even, then $I(\alpha, 1, 2) = \alpha$ and $I(\alpha, 2, 4) = (1 - cos2\alpha)/4$. If j is odd, $I(\alpha, 1, 3) = 1 - cos\alpha$ and $I(\alpha, 2, 4) = sin\alpha$.

2. **First recursion :** For $k = 4, \ldots, j$ if j is even ; for k=5,...,j if j is odd:

$$I(\alpha, k-1, k) = -\frac{(sin\alpha)^{k-3}cos\alpha}{(k-2)} + \frac{(k-3)}{(k-2)}I(\alpha, k-3, k-2)$$

For $k = 6, \ldots, j$ if j is even and for $k = 5, \ldots, j$ if j is odd:

$$I(\alpha, k-2, k) = -\frac{(sin\alpha)^{k-4}(cos\alpha)^2}{(k-2)} + \frac{(k-4)}{(k-2)}I(\alpha, k-4, k-2)$$

3. **Second recursion :** For $i = (j-3), (j-4), \ldots, 1$:

$$I(\alpha, i, j) = \frac{(sin\alpha)^i(cos\alpha)^{j-i-2}}{i} + \frac{(j-i-2)}{i}I(\alpha, i+2, j)$$

Appendix E: Nederlandse Samenvatting

Dit doktoraatswerk behandelt *wiskundige koncepten en algoritmen voor modellen van statische en dynamische systemen.*
De resultaten kunnen *grosso modo* in vier grote delen worden samengevat.

1. Niet-negatieve lineaire algebra en het veralgemeende lineaire komplementariteitsprobleem (hoofdstuk 2 en 3).

2. Koncepten en algoritmen in verband met georiënteerde energie en georiënteerde signaal-signaal verhoudingen (hoofdstuk 4).

3. Identifikatie van lineaire relaties voor statische systemen (hoofdstuk 5 en 6).

4. Identifikatie van dynamische lineaire modellen (hoofdstuk 7 en 8).

De verbanden tussen de verschillende hoofdstukken zijn weer te vinden in volgend schema:
De gebruikte techieken zijn terug te vinden in de *geometrische lineaire algebra* en de *numerieke wiskunde.* Zo wordt veelvuldig gebruikt gemaakt van vektorruimten (kolom- en rijruimten van matrices) en van daaraan gekoppelde begrippen zoals kanonieke hoeken tussen ruimten, nulruimtes, enz.... Ook *orthogonaliteit* staat centraal als basisbegrip, o.a. bij het identificeren van lineaire kombinaties tussen bepaalde observaties. Een en ander wordt zoveel mogelijk afgeleid en bestudeerd via technieken uit de numerieke wiskunde, die zich bij uitstek lenen tot een *numeriek betrouwbare en robuuste implementatie.* Zo wordt veelvuldig gebruik gemaakt van de (veralgemeende) singuliere waarden ontbinding. De op deze wijze geformuleerde eigenschappen en koncepten kunnen dan ook zondermeer numeriek worden uitgeprobeerd en geverifiëerd, waarbij bijna vanzelfsprekend aan de vereisten van numerieke stabiliteit voldaan wordt.

In wat volgt, worden de belangrijkste resultaten uit elk hoofdstuk samengevat. Hoofdstuk 1 bevat een algemene inleiding terwijl hoofdstuk 9 de algemene besluiten bevat.

Hoofdstuk 2: Niet-Negatieve Lineaire Algebra.

Niet-negatieve lineaire algebra is lineaire algebra waarbij aan sommige variabelen opgelegd wordt dat ze niet-negatief moeten zijn. Een en ander heeft een sterk geometrische inslag. In een eerste deel van hoofdstuk 2 wordt aandacht besteed aan de geometrische objecten die door dergelijke niet-negativiteitseisen geïnduceerd worden: (gepunte) polyhedrale kegels,

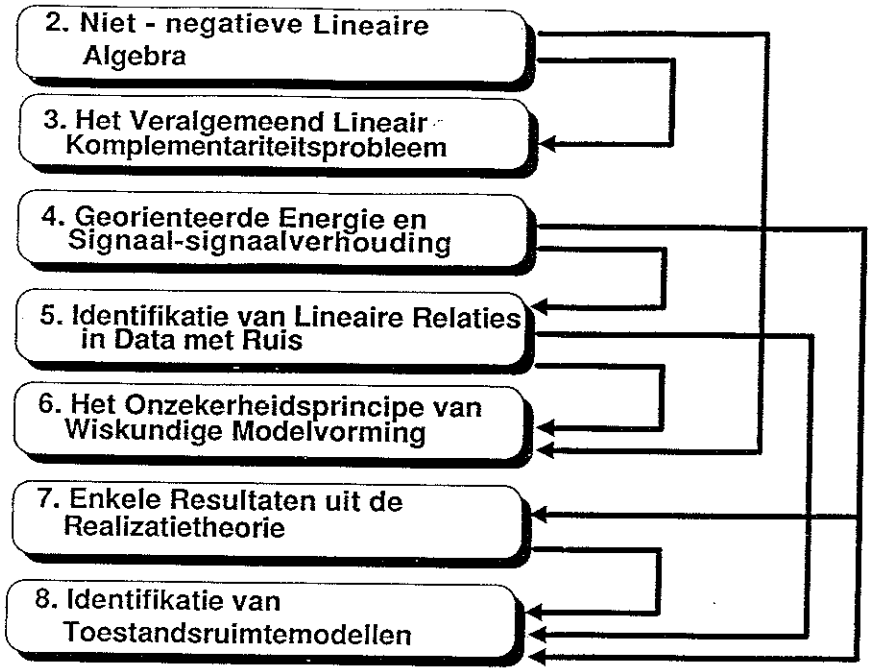


Figure 8.1: Verbanden tussen de verschillende hoofdstukken.

polyhedra, polytopen, simplices, enz.... Vervolgens wordt een algoritme afgeleid dat alle niet-negatieve oplossingen berekent voor een stel van homogene lineaire gelijkheden en ongelijkheden. De oplossingsverzameling van dit probleem is *een gepunte polyhedrale kegel* en het algoritme bepaalt *alle extreme ribben van deze polyhedrale kegel.* Het algoritme is rekursief in de vergelijkingen en de geldigheid ervan kan bewezen worden door middel van een dubbele induktieve bewijsvoering:

1. Vooreerst wordt voor elke bijkomende vergelijking, de reeds bekomen polyhedrale oplossingskegel aangepast, zodat hij alle nieuwe oplossingen bevat. Dit gebeurt door konvexe kombinaties te beschouwen van reeds gevonden ribben.

2. Vervolgens wordt bepaald welke oplossingen niet langer extreem zijn na de aanpassing. Hiertoe worden nodige en voldoende voorwaarden afgeleid die toelaten om te testen welke oplossingen extreem zijn en welke ribben van de polyhedrale kegel aangrenzend zijn.

Hoofdstuk 3: Het Veralgemeende Lineaire Komplementariteitsprobleem.

Het Veralgemeende Lineaire Komplementariteitsprobleem komt neer op het volgende matrix-vektor probleem:

Gegeven een $m \times n$ matrix Z. Vind alle vektoren u (of besluit dat er geen bestaan), die voldoen aan:

$$Zu = 0 \qquad u \geq 0 \qquad \sum \prod u_i = 0$$

Dit veralgemeend lineair komplementariteitsprobleem is een veralgemening van het 'klassieke' lineaire komplementariteitsprobleem in verschillende opzichten:

- De komplementariteitskondities zijn algemener dan de gangbare kombinatie van niet-negativiteit en orthogonaliteit. De komplementariteitskondities zijn in hun meest algemene vorm te schrijven als een som van produkten van komponenten van oplossingsvektoren.

- De matrix formulering van het probleem is algemener. Als een speciaal geval kan niet alleen het klassieke lineaire komplementariteitsprobleem met *singuliere* matrices opgelost worden, maar ook komplementariteitsproblemen die niet met de klassieke benadering kunnen worden opgelost. Deze algemenere matrix formulering houdt ook een verrijking in van het aantal modelleringsproblemen dat met deze techniek kan worden aangepakt.

Het algoritme om dit veralgemeende lineaire komplementariteitsprobleem op te lossen, is gebaseerd op een drie-dubbele induktieve bewijsvoering:

1. De eerste induktie toont aan dat het probleem in eerste instantie kan geformuleerd worden als het vinden van alle niet-negatieve oplossingen van een stelsel van linear vergelijkingen, waarbij de polyhedrale oplossingskegels rekursief (vergelijking per vergelijking) aangepast worden met het in hoofdstuk 2 bestudeerde algoritme.

2. De tweede induktie laat toe om in elke rekursiestap, alle redundante vektoren te elimineren, zodat alleen de extreme ribben van de polyhedrale oplossingskegel overblijven.

3. De derde induktie tenslotte laat toe om in elke rekursiestap, die extreme ribben te verwijderen, waarvan de komponenten niet voldoen aan de vereiste komplementariteitsvoorwaarden.

De belangrijkste eigenschappen van het algoritme zijn de volgende:

- Het bewijs van de geldigheid van het algoritme is konstruktief in die zin dat meteen ook wordt aangetoond dat de oplossingsverzameling van het veralgemeende lineaire komplementariteitsprobleem bestaat uit de unie van begrensde en onbegrensde polyhedra (diskrete vektoren, polytopen, polyhedral kegels).

- Het algoritme vindt *alle* oplossingen (in tegenstelling tot de klassieke algoritmen), zelfs indien er oneindig veel oplossingen zijn of indien er ook oplossingen *op oneinding* zijn.

- Door de zogenaamde *kruiskomplementariteiten* te verifiëren, is het mogelijk om na te gaan welke niet-negatieve kombinaties van oplossingen ook een oplossing zijn. Dit houdt in dat de oplossingsverzameling geometrisch volledig kan gekarakteriseerd worden.

- Er zijn geen matrixinversies of andere matrix berekenigen die eventueel slecht gekonditioneerd kunnen zijn, vereist. Dit verhoogt de numerieke betrouwbaarheid aanzienlijk in vergelijking tot klassiekere algoritmen, waar een dergelijke matrix operatie vereist is om het probleem in de juiste vorm te krijgen.

Een aanzienlijk aantal toepassingen en voorbeelden werden uitgewerkt:

- Er wordt aangetoond dat geometrische problemen met stuksgewijze lineaire relaties herleid kunnen worden tot een veralgemeend lineair komplementariteitsprobleem. Hiertoe werden verschillende parametrizaties voor stuksgewijze lineaire relaties in 2 en meerdere variabelen (dimensies) afgeleid, elk met hun specifieke voordelen.

- Er wordt aangetoond hoe het veralgemeende lineaire komplementariteitsprobleem toelaat om kwadratische optimizatievraagstukken op te lossen.

- Het veralgemeende lineaire komplementariteitsprobleem speelt een sleutelrol bij de studie van stuksgewijze lineaire weerstandsnetwerken. Er wordt aangetoond hoe de nieuwe techniek in staat is om *alle* oplossingen in deze toepassingen te vinden, daar waar de klassieke methodes slechts in staat waren om een beperkt deel van de oplossingsverzameling terug te vinden.

- Er wordt aangetoond hoe de berekening van *alle* invariante toestanden van een neuraal netwerk, met of zonder a priori partiële informatie, in feite niets anders is dan een veralgemeend lineair komplementariteitsprobleem.

Er kan echter nog gesleuteld worden aan de hoeveelheid rekenwerk die nodig is om alle oplossingen van een veralgemeend lineair komplementariteitsprobleem te berekenen. Dit rekenwerk is afhankelijk van twee parameters: Vooreerst volgt uit de aard van de oplossingsverzameling zelf dat het aantal extreme ribben van de polyhedrale oplossingskegel zeer groot kan zijn. Vervolgens heeft ook de volgorde waarin de verschillende vergelijkingen in de rekursie worden betrokken, een belangrijke invloed.

Hoofdstuk 4: Georiënteerde Energie en Georiënteerde Signaal-Signaal Verhoudingen in de Analyse van Vektorsekwenties en Tijdsreeksen.

In heel wat toepassingen, bestaan de metingen en de observaties uit één of meerdere vektorsekwenties. Deze worden typisch gegroepeerd in overgedetermineerde matrices met bijvoorbeeld veel meer kolommen n dan rijen m. De georiënteerde energie van zo'n vektorsekwentie is niets anders dan een maat voor de ruimtelijke verdeling van de vektoren van deze sekwentie in de n-dimensionale vektorruimte. Zo kunnen bepaalde richtingen sterker vertegenwoordigd zijn dan andere richtingen. In hoofdstuk 4 wordt aangetoond hoe de singuliere waardenontbinding een kanonieke ontbinding levert voor de georiënteerde energie. Zij laat toe om richtingen van sterkste en zwakste georiënteerde energie te berekenen, alsook tussengelegen *zadelpunten*. Wanneer de georiënteerde energie van een tweede vektorsekwentie dient te worden vergeleken met deze van een eerste sekwentie, dient men de georiënteerde signaal-signaal verhouding van twee vektorsekwenties te definiëren. Grosso modo komt dit koncept hier op neer, dat in elke richting van de n-dimensionale vektorruimte, de relatieve sterkte van de eerste vektorsekwentie wordt vergeleken met deze van de tweede. De veralgemeende singuliere waarden ontbinding laat toe om deze extreme richtingen van georiënteerde signaal-signaal verhoudingen te berekenen, d.w.z. de richtingen waar het eerste signaal het tweede maximaal overheerst in termen van georiënteerde energie en vice versa.
Een en ander wordt geïllustreerd met verschillende voorbeelden zoals Gauss-Markov minimum variantie schatters, de identifikatie en realizatie van dynamische systemen, de totale kleinste kwadratenbenadering, en enkele metoden die allemaal terug te brengen zijn op een of andere vorm van faktor-analyse: hoge resolutie signaal scheidingsalgoritmen en de scheiding van de moeder hartslag van deze van een foetus.

Hoofdstuk 5: Identifikatie van Lineaire Relaties in Data met Ruis

Er bestaat een nauw verband tussen *ortogonaliteit* (of beter de afwezigheid ervan) en de aanwezigheid van *lineaire relaties tussen bepaalde observaties.* Dit wordt toegelicht in de inleiding van hoofdstuk 5. Er wordt aangetoond hoe in elke identifikatiestrategie datgene wat niet door het model kan verklaard worden, als ruis dient te worden beschouwd. Een en ander leidt dan tot twee mogelijke instellingen wat betreft modelvorming: Een deduktieve benadering en een inspiratieve benadering. Een *deduktieve benadering* van het modelvormingsprobleem bestaat hoofdzakelijk uit een *perturbatie-analyse.* Er wordt verondersteld dat de 'echte, exakte' data door allerhande pertubaties niet exakt kunnen worden waargenomen. Deze perturbaties, die *op zich* ongekend zijn, worden gekenmerkt door globale beschrijvingen, zoals bijvoorbeeld waarschijnlijkheidsdichtheidsfunkties.
De *inspiratieve manier* van modelleren is in feite niets anders dan een probleem van *optimale benadering:* Hierbij wordt dat model gezocht uit een bepaalde modelklasse, dat het best de gegevens verklaard, volgens een of ander kriterium. Vanzelfsprekend zijn deze twee modelvormingsbenaderingen extremen en kunnen tussenliggende strategiën ook worden gedefiniëerd.
Voor beide instellingen echter, is het probleem van de *konsistentie* belangrijk: Een *deduktieve* modelvormingsstrategie is *konsistent* wanneer, onder bepaalde voorwaarden, het geschatte model konvergeert naar het 'exakte', wanneer het aantal observaties steeds maar toeneemt. In de *inspiratieve* aanpak, wordt een modelvormingsstrategie *konsistent* genoemd, indien voor 'exakte, ruisloze' data, het juiste, 'exakte' model wordt teruggevonden.

In hoofdstuk 5 worden twee belangrijke stellingen afgeleid en bewezen: het *ortogonaliteitstheorema* en de *hefboomstelling.*
In het ortogonaliteitstheorema worden de dichtheidsfunkties berekend die de waarschijnlijkheid uitdrukken dat vaste, exakte vektoren, een bepaalde hoek vormen met vektoren, waarvan de komponenten gegenereerd worden door een sferische waarschijnlijkheidsverdeling. Een belangrijke konklusie is dat, voor een toenemend aantal observaties, exakte vektoren met willekeurig gegenereerde vektoren, met toenemende waarschijnlijkheid hoeken gaan maken die meer en meer naar 90° neigen.
Dit resultaat wordt dan verder gebruikt in de hefboomstelling: Deze toont aan dat de zogenaamde *korte ruimte* van een matrix exakt kan worden teruggevonden wanneer de exakte data door additieve ruis verontreinigd zijn. De *lange ruimte* is echter onherroepelijk verloren. De hefboomstelling geeft een asymptotische uitdrukking voor de singuliere waarden ontbinding van een matrix, die door additieve ruis werd verontreinigd.

Deze twee stellingen spelen een sleutelrol bij de *analyse van klassieke identifikatieschema's*, zoals de (totale) kleinste kwadratenbenadering en bij het analyseren van de *asymptotische bias* op de schatting van de kanonieke hoeken tussen deelruimten, wanneer de gegevens niet exakt zijn. Verder geven deze stellingen ook aan wat de reden is van de zogenaamde *Mahalanobis transformatie*, waarbij een inprodukt gehanteerd wordt met als gewichtsmatrix de inverse van de Grammiaan van de kovariantiematrix van de ruis. Dit levert voor de *hefboomstelling* een belangrijke interpretatie op in termen van georiënteerde energie.

Ook in hoofdstuk 5 worden klassieke identifikatietechnieken met elkaar vergeleken, vooral

wat betreft hun *konceptuele* achtergrond. Er wordt aangetoond dat de techniek van de *lineaire kleinste kwadraten schatters* en de *totale lineaire kleinste kwadratenbenadering*, speciale gevallen zijn van rang 1 korrekties van de data matrix of van de Grammiaan (kovariantie) van de data matrix. Bovendien zijn alle kleinste kwadraten benaderingen van het probleem terug te vinden als de kolommen van de omgekeerde van de Grammiaan van de gegevensmatrix. Voor voldoende hoge signaal-ruis verhoudingen is de totale kleinste kwadratenbenadering te schrijven als een niet-negatieve lineaire kombinatie van alle kleinste kwadraten oplossingen. Er wordt ook aangetoond hoe het ruis model dat bij de totale kleinste kwadraten benadering hoort, eigenlijk een rang 1 matrix is. Dit is konceptueel gezien ontoelaatbaar. Er bestaat echter een meer realistisch ruismodel, dat de eenheidsmatrix heeft als kovariantie, maar waarvan de oplossing samenvalt met deze van de totale kleinste kwadratenbenadering. Dit toont aan dat de gekende konsistentie-eigenschap van de totale kleinste kwadratenbenadering eigenlijk bijna toeval is.

In een laatste deel van hoofdstuk 5, wordt onderzocht hoe *kanonieke hoeken tussen deelruimten* kunnen gedefiniëerd en berekend worden en hoe een *benaderende doorsnede van twee deelruimten* kan worden bepaald, indien de gegevens niet ruisvrij zijn. Een van de sleuteltechnieken is de veralgemeende singuliere waardenontbinding. Na een doorgedreven deduktieve analyse van de perturbatie van kanonieke hoeken onder de invloed van additieve ruis, worden verschillende optimizatiekriteria vooropgesteld om een benaderende doorsnede te berekenen. De enen leiden tot het bestuderen van de *kanonieke hoeken* terwijl de anderen gebaseerd zijn op de *veralgemeende singuliere waarden.* De techniek van kanonieke korrelatie is echter niet altijd even geschikt omdat het een expliciete normalisatiestap bevat, waarbij signaalmodes van zwakke energie, op gelijke voet worden behandeld als signaalmodes van sterke energie. Daartoe wordt een nieuw kriterium vooropgesteld dat de relatieve energiesterkten van de verschillende signalen impliciet in rekening brengt en dit in termen van veralgemeende singuliere waarden. Een en ander wordt geïllustreerd met simulaties.

Hoofdstuk 6: Het onzekerheidsprincipe van de wiskundige modelvorming.

In hoofdstuk 6 wordt een *onzekerheidsprincipe voor de wiskundige modelvorming* behandeld. In wezen betreft het een eenvoudig maar diepgaand inzicht: Wanneer de gegevens niet verzoenbaar zijn met modellen uit een bepaalde klasse, dan moet elk model dat uit die klasse gekozen wordt om de gegevens te 'verklaren', gekenmerkt worden door een welbepaalde onzekerheid. Deze onzekerheid weerspiegelt de onverzoenbaarheid van het model t.o.v. de gegevens. Terzelfdertijd houdt dit onzekerheidsprincipe in dat elke identifikatietechniek die een *uniek* model aflevert in het geval de gegevens niet ruisvrij zijn, beroep moet doen op veronderstellingen die niet a priori uit de gegevens zelf voortvloeien. Alleen op deze manier kan de *uniciteit* van het model, gegarandeerd worden.

In hoofdstuk 6 wordt aangetoond hoe de *identifikatie van lineaire relaties* uit *niet-ruisvrije gegevens*, wiskundig kan herleid worden tot de minimizatie van de rang van een positief definiete matrix, waarvan alleen de diagonale elementen mogen gewijzigd worden zodat het resultaat toch nog niet-negatief definiet is. Eerst wordt een historisch overzicht gegeven van vooral *algebraische* benaderingen tot oplossing van dit probleem, zoals bepaalde resultaten over de-

terminanten of pogingen waarin gebruik wordt gemaakt van het Perron-Frobeniustheorema voor elementsgewijze positieve matrices.

In de door ons gevolgde werkwijze, wordt een meer *geometrische* weg bewandeld. Er wordt aangetoond hoe de oplossingsvektoren, die de benaderende lineaire relaties beschrijven, aan drie nodige voorwaarden moeten voldoen: zij moeten *orthant invariant* zijn, *nul-invariant* en ook *toegelaten.* Er wordt ook aangetoond dat, wanneer de minimale rang die kan bereikt worden r is, er minstens $n-r$ lineair onafhankelijke toegelaten orthant nul invariante oplossingen bestaan, elk met minstens $n-r-1$ nullen (n is hier het aantal variabelen). Het algoritme van hoodstuk 2 voor het bepalen van alle niet-negatieve oplossingen van een stelsel van lineaire vergelijkingen, kan gebruikt worden om alle toegelaten orthant nul invariante vektoren met nullen te vinden. Dit laat toe om numeriek op een geometrische geïnspireerde manier, de minimal rang van het probleem te vinden.
Tenslotte wordt ook de genericiteit van de oplosbaarheid behandeld. Er wordt uitgelegd hoe er een ondergrens bestaat op de generisch bereikbare minimal rang als een funktie van het aantal variabelen. Dit is de zogenaamde Wilson-Ledermann ondergrens.

Het volledige bewijs van de geldigheid van onze oplossingsmanier is echter nog niet geleverd: Er zijn vele aanwijzingen die ons in ons vermoeden sterken dat de oplossingsverzameling bestaat uit een verzameling van polyhedrale kegels van toegelaten orthant nul invariante vektoren. De ribben van deze kegels zouden dan de door ons algoritme gevonden orthant nul invariante vektoren met nullen zijn, samen met de klassieke kleinste kwadraten oplossingen. Alle vektoren uit deze kegels voldoen aan de vereisten van het identifikatieschema. Bijgevolg zijn deze kegels de rechtstreekse uiting van de onverzoenbaarheid van de gegevens met de veronderstelde lineariteit van het model.

Hoofdstuk 7: Enkele Resultaten uit Realizatietheorie.

In hoofdstuk 7 wordt het verband uitgediept dat bestaat tussen de *shiftstruktuur* van bepaalde deelruimten en lineaire tijdsinvariante tijdsdiskrete systemen. De singuliere waardenontbinding van de blokhankelmatrix van Markovparameters van een systeem geeft rechtstreeks aanleiding tot verschillende realizatiealgoritmen, levert bovendien kwantitatieve maten voor kontroleer- en observeerbaarheid en biedt een verklaring voor de goede werking van modelreduktiestrategieën in termen van georiënteerde energie.
De shiftstruktuur van kolom- en rijruimten van een rangdeficiënte blokhankelmatrix komt tevoorschijn via om het even welke faktorisatie van deze blokhankelmatrix, zoals bijvoorbeeld de singuliere waarden ontbinding. Wanneer gebruik gemaakt wordt van de niet-konventionele matrixprodukten zoals Kronecker en Khatri-Rao produkten, kan worden aangetoond dat alle rangdeficiënte blokhankelmatrices van bepaalde (eindige) dimensies, die een bepaalde verzameling van minimale systeem polen delen, kunnen geparemetrizeerd worden via de nulruimte van een bepaald matrixprodukt.
Deze resultaten laten toe om de eigenschappen van orthonormale blokhankelmatrices te bestuderen en om de shift struktuur te veralgemenen naar matrices zonder specifieke struktuur maar met een *impliciete* shiftstruktuur. Zo kan men in een bepaalde deelruimte, de shiftinvariante ruimte van maximale dimensie bepalen. Dit suggereert dan verschillende toepassingen in de aard van faktor-analyse, die eerder de impliciete shiftstruktuur uitbuiten dan de relatieve

signaal-signaal sterkte.

Hoofdstuk 8: Identifikatie van Toestandsruimtemodellen.

In hoofdstuk 8 worden vooreerst enkele belangrijke, doch minder of niet gekende eigenschappen van exakte toestandsruimtemodellen afgeleid en aangetoond. Zo wordt bijvoorbeeld bewezen hoe de blokhankelmatrix van uitgangsvektoren kan geschreven worden als een som van produkten: het produkt van een blokhankelmatrix met Markovparameters en een bloktoeplitzmatrix met ingangsvektoren en het produkt van een bloktoeplitzmatrix met Markovparameters en een blokhankelmatrix met ingangsvektoren. Vervolgens wordt een belangrijke rangeigenschap bewezen: de rang van de gekombineerde ingangs-uitgangs blokhankelmatrix is een som van drie termen: het geobserveerde geëxciteerde deel van de toestandsvariabelenruimte, de rang van de ingangsblokhankelmatrix verminderd met de doorsnede van de rijruimte van de toestandssekwentie met deze van de ingangsblokhankelmatrix. Deze belangrijke rangeigenschap wordt in detail geanalyseerd. Vervolgens wordt aangetoond hoe het probleem van kausale afhankelijk kan aangepakt worden. Hierbij wordt aangetoond hoe in principe wiskundig kan bepaald worden wat een ingangsvariabele en wat een uitgangsvariabele van een systeem is. Ook wordt bondig behandeld wat de invloed is van een looptijd op de singuliere waarden van een ingangs-uitgangs blokhankelmatrix.

Gebaseerd op de verkregen inzichten worden dan verschillende identifikatiestrategieën bestudeerd. Zij zijn allen gebaseerd op een welbepaald geometrisch idee dat dan vertaald wordt in een numeriek betrouwbaar algoritme dat typisch gebaseerd is op de (veralgemeende) singuliere waardenontbinding. Zo worden een heuristische lineair en totale lineaire kwadratische identifikatiemethode behandeld, waarbij aandacht wordt besteed aan een deduktieve perturbatieanalyse in het geval van data met additieve ruis. Andere identifikatie methoden volgen uit de belangrijke stelling dat de toestandssekwentie kan bepaald worden uit de doorsnede van de rijruimten van verleden en toekomstige ingangs-uitgangs blokhankelmatrices. Ook hier worden verschillende versies behandeld: (totale) kleinste kwadraten, voorwaartse en achterwaartse prediktie, allen gebaseerd op kanonieke korrelatietechnieken en ook een niet-genormalizeerde berekening van de doorsnede, gebaseerd op veralgemeende singuliere waarden. De systeemmatrices worden dan berekend als de oplossing van een stelsel lineaire vergelijkingen. Een en ander kan geoptimaliseerd worden wat betreft de benodigde rekenintensiteit door een rekonversie naar de zogenaamde *korte* ruimte, wat dan weer uitstekend kan vertaald worden naar een on-line algoritme.

Tenslotte worden de verschillende theoretische inzichten geïllustreerd met de identifikatie van vier verschillende industriële toepassingen.

Appendix A : Polyhedrons, Polyhedral cones, Polytopes, Simplices

Most of the definitions and properties that play a crucial role in nonnegative linear algebra , are briefly summarized in this section and discussed in a rather informal way without proofs. For more details and rigor, we refer to excellent books, references of which can be found in the bibliography of chapter 2.

All scalars, vectors and matrices are real (or integer, if specified). The n-dimensional Euclidean space of real numbers is denoted by $\mathcal{R}^n$.

Affine and convex sets

Definition 1 Affine set
A subset S of the real d-dimensional Euclidean space $\mathcal{R}^d$ is called an affine set if it contains the line passing through any two of its distinct points. If $x, y \in S$, then $\gamma x + (1-\gamma)y \in S$ for all $\gamma \in \mathcal{R}$.

Definition 2 Affine combination
The linear combination $\sum_{i=1}^{m} x^i \gamma_i$ of vectors $x^i, i = 1, \ldots, m$ is called an affine combination if $\sum_{i=1}^{m} \gamma_i = 1$ for all $\gamma_i \in \mathcal{R}$.

A finite set of points is *linearly (affinely) independent* if none of its points can be expressed as a linear (affine) combination of the others. Consider k points in $\mathcal{R}^n$: $\{x^1, \ldots, x^k\}$ with coordinates x_i^j. The maximum number of linear independent points in this set is equal to the rank of the $n \times k$ matrix x_i^j. A linear space (affine set) is called *n-dimensional* if the maximal number of linearly (affinely) independent points in it is n $(n+1)$. Let S be an arbitrary non-empty subset of $\mathcal{R}^n$. Then the set of all affine combinations of points taken from S is an affine set called the *affine hull* of S : aff(S).

Definition 3 Nonnegative linear combinations
A nonnegative linear combination of a set of vectors x^i, $i = 1, \ldots, q$ is any linear combination of the form $\sum_{i=1}^{q} x^i \alpha_i$ where coefficients $\alpha_i, i = 1, \ldots, q$ are real nonnegative *numbers.*

Definition 4 Convex combinations
A convex combination of a set of vectors $x^i, i = 1, \ldots, q$ is a nonnegative combination $\sum_{i=1}^{q} x^i \alpha_i$ with the additional constraint that $\sum_{i=1}^{q} \alpha_i = 1$

Let W be a non-empty set in $\mathcal{R}^n$.

Definition 5 Convex sets
The line segment joining the points $x, y \in \mathcal{R}^n$ consists of the set of points $x\gamma + y(1-\gamma)$, $0 \leq \gamma \leq 1$, The set $W \subset \mathcal{R}^n$ is called a convex set if it contains the line segment joining any of its points.

The dimension of a convex set $W \subset \mathcal{R}^n$ is defined to be the dimension of its affine hull.

Theorem 1 *The intersection of any number of convex sets is a convex set.*

Definition 6 Convex hull
The set of all convex combinations of points taken from W is a convex set called the convex hull of W, denoted conv(W).

A classical result, due to Caratheodory, shows that in constructing the convex hull of $W \in \mathcal{R}^n$, it is not necessary to take combinations of more than $n+1$ points:

Theorem 2 *The convex hull of a set $W \subset \mathcal{R}^n$ is the union of all convex combinations of all subsets of W containing no more than $n+1$ points.*

Another characterization of convex sets goes via its extreme vectors.

Definition 7 Extreme vectors
A vector x of a convex set W is an extreme point if it is not an interior point of any line segment with distinct endpoints in W.

Definition 8 Bounded sets
A set is called bounded if it is contained within some sphere.

Theorem 3 *A non-empty closed bounded convex set in $\mathcal{R}^n$ possesses extreme points and is the convex hull of the set of its extreme points.*

Let W be a non-empty set in $\mathcal{R}^n$.

Definition 9 Supporting hyperplane
The hyperplane H is called a supporting hyperplane to the set W if H has at least one point in common with W and if W is contained in one of the two closed half-spaces H^+ and H^- defined by H.

The half-space containing W is called a supporting half-space to W.

Linear inequalities and polyhedrons

One of the cornerstones of nonnegative linear algebra, is the study of systems of linear inequalities, started by mathematicians such as Minkowski, Weyl and others.

The definitions of the preceding section already provide a lot of geometrical insight into the problem of solving nonnegatively a set of linear inequalities. Suppose that the vectors x_1 and x_2 are both nonnegative solutions to the problem $Ax \geq b$ and assume that b is not identically 0. Then clearly a nonnegative combination $x_1\alpha_1 + x_2\alpha_2$ will be a solution too provided that $b\alpha_1 + b\alpha_2 = b$, which is equivalent with $\alpha_1 + \alpha_2 = 1$. Hence, any *convex* combination of two solutions is also a solution ! Hence the solution set is convex ! If $b = 0$, then it is easy verified that any *nonnegative* linear combination of two solutions is also a solution. These simple facts are now rigorously stated in the following theorems and properties.

Theorem 4 *The set of solution points of any (finite or infinite) system of linear inequalities* $a_i^t x \leq b_i, i = 1, \ldots..$ *is either convex or empty (if the system is inconsistent).*

As an example of a convex set which is the solution to an infinite system of linear inequalities, consider in $\mathcal{R}^2$ the closed disc of radius r, defined by:

$$xcos(\delta) + ysin(\delta) \leq r \quad 0 \leq \delta \leq 2\pi \quad \delta \in \mathcal{R}$$

In our work, we are mainly interested in the solution(s) of a finite number of linear inequalities.

Definition 10 Polyhedron
The solution of a finite system of linear inequalities is called a polyhedron.

Polyhedral Cones

Definition 11 Cones
A subset $K \subset \mathcal{R}^n$ *is called a cone if* $\gamma x \in K \ , \forall x \in K$ *and* $\gamma \geq 0$.

Definition 12 Convex Cones
A convex cone is a cone which is also a convex set.

As a special result, note that any linear space in $\mathcal{R}^n$ is a convex cone. Another example is a halfspace in $\mathcal{R}^n$ defined by a hyperplane passing through the origin.

Theorem 5 Intersection of convex cones
The intersection of convex cones is a convex cone.

This implies that the set of solutions to a *finite system of homogeneous linear inequalities is a convex cone .*

Definition 13 Polyhedral Cone
A polyhedral cone is the set of solutions of a finite system of homogeneous linear inequalities.

Definition 14 Conical Combination
A vector $x = \sum_{i=1}^{m} x_i\gamma_i \ , \ \gamma_i \geq 0 \ , \ i = 1, \ldots, m$ *is called a conical combination of the points* $x_1, \ldots, x_m$.

Let S be a non-empty set in $\mathcal{R}^n$. The set con(S) of conical combinations of all subsets of S is a convex cone called the cone generated by S. A convex cone generated by a finite set of vectors is called multifaceted. The following fundamental theorem in the theory of convex cones and the theory of linear inequalities, is due to Weyl:

Theorem 6 *A convex cone K is polyhedral if and only if it is multifaceted.*

This result means that a polyhedral cone is finitely generated and can be represented as a (finite) enumeration of its extreme rays !

Polytopes

Definition 15 Polytope
The convex hull of a finite sets V of points in $\mathcal{R}^n$ is called the polytope generated by the points of V.

Let H be a supporting hyperplane of the polytope M.

Definition 16 Faces
The set $F = M \cap H$ is a face of the polytope M, generated by H. If $dim(F) = i$, then F is an i-face of the polytope M. The 0-faces are called the vertices of M. 1-faces are called edges. The empty set $\emptyset$ and M itself are called improper faces. All other faces are proper faces. If $dim(M) = d$, then the $(d-1)$-faces of M are called facets of M. They are the proper faces of maximal dimension.

The face structure of a polytope in $\mathcal{R}^3$ is quite easy to visualize. However, in the higher dimensional Euclidean space, things are not that straightforward. However, we 'd like to mention a very celebrated (and beautiful) formula relating the number of faces of each dimension of a d-polytope M, which was derived in its general form by Poincare.

Theorem 7 The Euler-Poincare formula
Let M be a n-polytope. Then:

$$\sum_{i=0}^{n-1}(-1)^i F_i(M) = 1 + (-1)^{n-1}$$

where $F_i(M)$ is the number of i-faces of M.

For $d = 2$, this result specializes to the celebrate Euler formula, relating the vertices, edges and 2-faces of a 3-polytope:

$$f_0 - f_1 + f_2 = 2$$

The Euler-Poincare formula establishes a linear dependence between the number of faces of the different dimensions of any d-polytope. It is remarkable that it can also be proven that there are no other linear relations that describe this linear dependence for an arbitrary polytope of fixed dimension (i.e. generically), although for some polytopes with a definite structure other linear relations may apply. The following theorems provide some additional useful insights.

Theorem 8 *A polytope has a finite number of distinct faces and each of its faces is itself a polytope.*

Theorem 9 *A polytope is the convex hull of its vertices.*

Theorem 10 *The vertices of a polytope are its only extreme points.*

The preceding theorems allow us to give an equivalent characterization of the vertices of a polytope.

Definition 17 *A point of a polytope M is a vertex if it cannot be represented as a convex combination of any other two distinct points of M.*

There is a very precise relation between polytopes and polyhedrons, which was derived for the first time by Minkowski. and which is essential in describing the intimate relation between systems of linear inequalities and polytopes.

Theorem 11 The Weyl-Minkowski theorem
The set M is a polytope if and only if M is a bounded polyhedron.

Simplices

The simplest type of polytopes are simplices.

Definition 18 Simplex
The convex hull of an affinely independent set of points is a simplex.

In a n-dimensional vectorspace, an n-simplex is a polytope bounded by $(n+1)$ hyperplanes. A simplex in 2 dimensions is a triangle, in 3 dimensions it is a tetrahedron. Simplices have a rather transparant structure in that their number of faces can be counted explicitely:

Theorem 12 *Let $0 \leq k \leq n-1$. Every $(k+1)$-subset of the vertices of an n-simplex T_n determines a k-face. The number $f_k(T_n)$ of k-faces of a simplex T_n equals:*

$$\begin{pmatrix} n+1 \\ k+1 \end{pmatrix}$$

From all n-polytopes, the n-simplex has the minimal number of facets, namely $n+1$.