Notes for Andrew Young

A. P. Mullhaupt

October 22, 2010

Abstract.

1. Prediction of a Stationary Process

We are interested in predicting a series of numbers $y(t)$ in the case where such prediction is difficult, but not impossible. We restrict our attention to the case where t has integer values, which is the case of practical interest.

For simplicity, we consider a case which is special in several ways, but, which includes important features. In particular, we assume that for each t, $y(t)$ is a random variable which we are able to observe, this is essentially the same as assuming that we will not predict $y(t)$ itself, but we will predict the distribution from which $y(t)$ is drawn conditioned on the information from time . We assume that $y(t)$ is a stationary process with finite mean and covariance.

In view of Wold's theorem, we can represent the $y(t)$ in the form

$$y(t) = \delta(t) + \sum_{k=0}^{\infty} h(k)\, x(t-k) \tag{1}$$

where $\delta(t)$ is deterministic (which means $\delta(t) = E(\delta(t)|t-1)$), $h(0) = I$, $\sum_{k=0}^{\infty} \|h(k)\|_2^2 < \infty$, and $x(t) = y(t) - E(y(t)|t-1)$. $x(t)$ defined this way are called *innovations*, and the corresponding representation is called the innovations representation.

It follows that

$$E(y(t)|t-1) = \delta(t) + \sum_{k=1}^{\infty} h(k)\, x(t-k). \tag{2}$$

The implication is that the expectation of the nondeterministic part of a stationary process is a linear function of the history of innovations, that function given by the *impulse response* h.

One might be forgiven for thinking that there is little left to do at this point. In fact there are still considerable difficulties. We do not know the values of δ and h. Finding those values is called *system identification*. Once we have those values, we do not know the values $x(t)$.

However we do know one thing about the sequence $x(t)$. Since

$$x(t) = y(t) - E(y(t)|t-1) \tag{3}$$

we must have

$$E(x(t)|t-1) = E(y(t) - E(y(t)|t-1)|t-1) = 0 \tag{4}$$

and

$$E(x(t)) = E(y(t) - E(y(t)|t-1)) = 0. \tag{5}$$

Thus the innovations is a sequence of independent mean zero random variables.

We can then rewrite our decomposition as

$$\begin{pmatrix} E(y(t)|t-1) \\ E(y(t+1)|t-1) \\ E(y(t+2)|t-1) \\ \vdots \end{pmatrix} = \begin{pmatrix} \delta(t) \\ \delta(t+1) \\ \delta(t+2) \\ \vdots \end{pmatrix} + \begin{pmatrix} h(1) & h(2) & h(3) & \cdots \\ h(2) & h(3) & & \\ h(3) & & & \\ \vdots & & & \end{pmatrix} \begin{pmatrix} x(t-1) \\ x(t-2) \\ x(t-3) \\ \vdots \end{pmatrix}. \tag{6}$$

The block Hankel structure of the matrix is due to stationarity.

2. State Space

Here we rewrite the innovations representation in a useful form. Define vectors $z(t)$ according to

$$z(t) = \begin{pmatrix} x(t-1) \\ x(t-2) \\ x(t-3) \\ \vdots \end{pmatrix} \tag{7}$$

and observe that $z(t)$ satisfies the recursion

$$z(t+1) = \begin{pmatrix} 0 & & & \\ I & 0 & & \\ & I & 0 & \\ & & \ddots & \ddots \end{pmatrix} z(t) + \begin{pmatrix} I \\ 0 \\ 0 \\ \vdots \end{pmatrix} x(t) \tag{8}$$

which allows us to express the innovations representation

$$y(t) = \begin{pmatrix} h(1) & h(2) & h(3) & \cdots \end{pmatrix} z(t) + x(t). \tag{9}$$

These equations are in the 'state space' form

$$z(t+1) = Az(t) + Bx(t) \tag{10}$$

$$y(t) = Cz(t) + x(t) \tag{11}$$

where

$$A = \begin{pmatrix} 0 & & & \\ I & 0 & & \\ & I & 0 & \\ & & \ddots & \ddots \end{pmatrix} \tag{12}$$

$$B = \begin{pmatrix} I \\ 0 \\ 0 \\ \vdots \end{pmatrix} \tag{13}$$

$$C = \begin{pmatrix} h(1) & h(2) & h(3) & \cdots \end{pmatrix}. \tag{14}$$

Given the generality of the Wold decomposition, this can be taken to be the "grand unified theory" of linear time invariant systems.

2.1. Beware of imitations?. There are other similar ways of writing state space systems. For example it is pretty easy to argue that if $z(t)$ is a vector in the intersection of the span of $(x(t-1), x(t-2), ...)$ and $(y(t), y(t+1), ...)$ and the process is stationary, then you arrive at equations

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) \quad (15) \\ y(t) &= Cz(t) + Dx(t) \quad (16) \end{aligned}$$

where (A, B, C, D) are conformable matrices, but it is not obvious that you can choose $D = I$, as in the innovations representation.Well, you can, this is a theorem proved in Hannan and Deistler for example

There is also the approach from physical analogy. The idea here is that $z(t)$ is the 'state' of a physical system which is enough data for the evolution of the statistics of the system to be determined. Then $y(t)$ is an observation. Because the system does not evolve deterministically, and the observation has noise, then you arrive at these state space equations

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) \quad (17) \\ y(t) &= Cz(t) + Dw(t) \quad (18) \end{aligned}$$

where $w(t)$ and $x(t)$ are independent of values from previous times, but can be correlated with each other at time t. It is not obvious that you can find a representation where $w(t) = x(t)$; but as long as your system is not pathological (in what sense I will say later) then you can.

This section is here mainly to allow the reader to avoid problems reading other treatments. I draw attention to the fact that we are concerned with the innovations representation.

In the innovations representation, there are two cases: when $y(t)$ is a scalar, which is called the SISO (single input/single output) case, and when $y(t)$ is a vector, which is called the MIMO (multiple input/multiple output) case.

3. Factoring the Hankel Matrix.

The state space system

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) \quad (19) \\ y(t) &= Cz(t) + x(t) \quad (20) \end{aligned}$$

effects a convolution product of x and h, where h is the impulse response

$$h(k) = \begin{cases} 0 & k < 0 \\ I & k = 0 \\ CA^{k-1}B & k > 0 \end{cases}. \quad (21)$$

The property that $h(k) = 0$ for $k < 0$ corresponds to *causality*. The Wold decomposition guarantees that the impulse response of a stationary process satisfies $\sum_{k=0}^{\infty} \|h(k)\|_2^2 < \infty$.

The state space system is an example of a factorization of the block Hankel matrix

$$H = \begin{pmatrix} h(1) & h(2) & h(3) & \cdots \\ h(2) & h(3) & & \\ h(3) & & & \\ \vdots & & & \end{pmatrix} = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}. \quad (22)$$

In the innovations representation, the blocks $h(k)$ are square matrices. In the following we will call this sort of matrix a Hankel matrix, with the understanding that it is block Hankel.

Any full rank factors of a Hankel matrix must be of these *Krylov* forms. There is nothing preventing us from assigning C and B to the initial rows and columns of the factors, respectively. The key is to see that there is an A which completes the picture. Suppose

$$H = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \end{pmatrix} \begin{pmatrix} B_1 & B_2 & B_3 & \cdots \end{pmatrix} \tag{23}$$

is a block Hankel matrix, where we have partitioned the factors conformably.

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \end{pmatrix} \begin{pmatrix} B_1 & B_2 & B_3 & \cdots \end{pmatrix} = \begin{pmatrix} C_1B_1 & C_1B_2 & C_1B_3 & \\ C_2B_1 & C_2B_2 & & \\ C_3B_1 & & & \\ & & & \end{pmatrix} \tag{24}$$

so we have the constraints

$$C_k \begin{pmatrix} B_1 & B_2 & \cdots \end{pmatrix} = C_{k-1} \begin{pmatrix} B_2 & B_3 & \cdots \end{pmatrix} \tag{25}$$

$$\begin{pmatrix} C_1 \\ C_2 \\ \vdots \end{pmatrix} B_j = \begin{pmatrix} C_2 \\ C_3 \\ \vdots \end{pmatrix} B_{j-1} \tag{26}$$

which imply

$$C_k = C_{k-1}A_C \tag{27}$$

$$B_j = A_B B_{j-1} \tag{28}$$

where

$$A_C = \begin{pmatrix} B_2 & B_3 & \cdots \end{pmatrix} \begin{pmatrix} B_1 & B_2 & \cdots \end{pmatrix}^+ \tag{29}$$

$$A_B = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \end{pmatrix}^+ \begin{pmatrix} C_2 \\ C_3 \\ \vdots \end{pmatrix} \tag{30}$$

and M^+ denotes the Moore-Penrose pseudoinverse of M.

This establishes that the factors are Krylov matrices. We return to the Hankel factorization

$$H = \begin{pmatrix} C \\ CA_C \\ CA_C^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B & A_B B & A_B^2 B & \cdots \end{pmatrix} \tag{31}$$

and see that we have $CA_C^k B = CA_B^k B$ for all $k \geq 0$, so we also have

$$\begin{pmatrix} C \\ CA_C \\ CA_C^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B & A_B B & A_B^2 B & \cdots \end{pmatrix} = \begin{pmatrix} C \\ CA_C \\ CA_C^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B & A_C B & A_C^2 B & \cdots \end{pmatrix} \tag{32}$$

and by full rank of the factors

$$\begin{pmatrix} B & A_B B & A_B^2 B & \cdots \end{pmatrix} = \begin{pmatrix} B & A_C B & A_C^2 B & \cdots \end{pmatrix}. \tag{33}$$

Therefore

$$\begin{aligned} A_B &= A_B \begin{pmatrix} B & A_B B & \cdots \end{pmatrix} \begin{pmatrix} B & A_B B & \cdots \end{pmatrix}^+ && (34)\\ &= \begin{pmatrix} A_B B & A_B^2 B & \cdots \end{pmatrix} \begin{pmatrix} B & A_B B & \cdots \end{pmatrix}^+ && (35)\\ &= \begin{pmatrix} A_C B & A_C^2 B & \cdots \end{pmatrix} \begin{pmatrix} B & A_C B & \cdots \end{pmatrix}^+ && (36)\\ &= A_C \begin{pmatrix} B & A_C B & \cdots \end{pmatrix} \begin{pmatrix} B & A_C B & \cdots \end{pmatrix}^+ && (37)\\ &= A_C && (38) \end{aligned}$$

which completes the result.

If the Hankel factors are not full rank, then they need not be in the form of Krylov matrices. For example for any sequences $x_1, x_2, ...$ and $y_1, y_2, ...$ we have

$$\begin{pmatrix} C \\ CA \\ \vdots \end{pmatrix} \begin{pmatrix} B & AB & \cdots \end{pmatrix} = \begin{pmatrix} C & x_1 & x_1 \\ CA & x_2 & x_2 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} B & AB & \cdots \\ y_1 & y_2 & \cdots \\ -y_1 & -y_2 & \cdots \end{pmatrix}. \tag{39}$$

As this example illustrates, rank deficient factorizations have components which are unrelated to the Hankel matrix.

We can always replace a rank deficient factorization with a full rank factorization - suppose $H = FG$ where a column of F is in the span of the other columns of F - without loss assume that it is the first column:

$$\begin{aligned} F &= \begin{pmatrix} f_1 & f_2 & f_3 & \cdots \end{pmatrix} && (40)\\ &= \begin{pmatrix} \sum_{k\geq 2} c_k f_k & f_2 & f_3 & \cdots \end{pmatrix} && (41) \end{aligned}$$

then since

$$\begin{aligned} FG &= \begin{pmatrix} f_1 & f_2 & f_3 & \cdots \end{pmatrix} \left(I - \sum_{k\geq 2} c_k e_k e_1^T \right) \left(I + \sum_{k\geq 2} c_k e_k e_1^T \right) G && (42)\\ &= \left[\begin{pmatrix} 0 & f_2 & f_3 & \cdots \end{pmatrix} \right] \left[\left(I + \sum_{k\geq 2} c_k e_k e_1^T \right) G \right] && (43) \end{aligned}$$

we have a factorization where we can replace a dependent column of F by a column of zeros. The factorization is unchanged if we then delete that column from F, and the

corresponding row from $\left(I + \sum_{k\geq 2} c_k e_k e_1^T\right) G$. Removing a dependent row from G is entirely similar.

From the preceding, it follows that every Hankel matrix has a full rank factorization in Krylov matrices, which corresponds to a state space system. Such a full rank factorization is called *minimal*.

In most of what follows, we will be interested in cases where a Hankel matrix has a minimal factorization of finite rank. In a finite rank minimal factorization $FG = H$, then $\operatorname{rank} F = \operatorname{rank} G = \operatorname{rank} H$. (Since F is not rank deficient, then $F^T F$ is nonsingular, so $G = \left(F^T F\right)^{-1} F^T H$ and $\operatorname{rank} G \leq \min(\operatorname{rank} F, \operatorname{rank} H)$. But $\operatorname{rank} H \leq \operatorname{rank} G$. So we have $\operatorname{rank} H = \operatorname{rank} G$ and $\operatorname{rank} G \leq \operatorname{rank} F$. and similarly for $\operatorname{rank} F \leq \operatorname{rank} G$.)

The rank of a minimal factorization is also the dimension of the 'state' vectors $z(t)$ in the state space system corresponding to the factorization.

Remark 1. *The simplicity of removing dependent columns and rows from a Hankel factorization stands in stark contrast to the complexity of dealing intelligently with "nearly" dependent columns and rows. Dealing with exactly dependent rows and columns is* topological*, but dealing with "nearly" dependent columns and rows is* geometric*.*

4. Elementary Topology of State Space Systems

Here we revisit the concepts of the previous section, but with a different point of view which allows us a little more flexibility, and also allows us to introduce important concepts. In this section we have in mind the state space system

$$z(t+1) = Az(t) + Bx(t) \tag{44}$$

$$y(t) = Cz(t) + x(t) \tag{45}$$

where the dimension of the state space is finite, say n.

By the Cayley-Hamilton theorem, there is a polynomial p of degree less than or equal to n such that $p(A) = 0$. We normally write this polynomial as

$$\prod_{k=1}^{n} (\lambda_k - A) \tag{46}$$

where $\{\lambda_k\}$ are the eigenvalues of A (replicated for multiplicity). In particular this allows us to express A^n as a linear combination of $I, A, ..., A^{n-1}$. In view of this, the (semi-infinite) Krylov matrix

$$\begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \tag{47}$$

has columns spanned by the columns of the *reachability* matrix

$$\begin{pmatrix} B & AB & \cdots & A^{n-1}B \end{pmatrix}. \tag{48}$$

The Krylov matrix is full rank if an only if the reachability matrix is nonsingular. A state space system with this property is called reachable. There is a closely related concept of *controllability*, which is what reachability is called when the sequence of innovations $x(t)$ are considered as a sequence of *control inputs*. We also have the reachability (controllability) *Grammian*:

$$P = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ B^*A^{*2} \\ \vdots \end{pmatrix}. \tag{49}$$

The reachability Grammian is positive definite if and only if the reachability matrix is nonsingular, and the reachability Krylov matrix has full rank. The reachability Grammian legitimately is a Gram matrix - the elements are the inner products of the rows of the Krylov matrix with each other.

In what follows we we normally abuse the terminology and use controllability instead of reachability as the name for this property and it's Grammian.

The dual concept of observability refers to full rank of the Krylov matrix

$$\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix}, \tag{50}$$

nonsingularity of the observability matrix

$$\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} \tag{51}$$

and positive definiteness of the observability Grammian

$$Q = \begin{pmatrix} C^* & A^*C^* & A^{*2}C^* & \cdots \end{pmatrix} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix}. \tag{52}$$

It follows that a state space system is minimal if and only if it is observable and controllable, thereby providing a minimal factorization of the corresponding Hankel matrix.

5. Composition of State Space Systems

A state space system effects a convolution product of the impulse response h and the input x. It turns out that deconvolving the output y to recover x is always possible in an innovations representation. Explicitly:

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) && (53)\\ y(t) &= Cz(t) + x(t) && (54) \end{aligned}$$

can be simply rearranged to form the *inverse* system

$$\begin{aligned} z(t+1) &= (A - BC)z(t) + By(t) && (55)\\ x(t) &= -Cz(t) + y(t) && (56) \end{aligned}$$

where y now appears as the input, and x is the output. The system is called the inverse because it transforms y back into x One sees that composition of state space systems is defined by identifying the output of one state space system with the input of another, so that the impulse responses are convolved with each other:

$$\begin{aligned} z_1(t+1) &= A_1 z_1(t) + B_1 x(t) && (57)\\ \xi(t) &= C_1 z_1(t) + x(t) && (58)\\ z_2(t+1) &= A_2 z_2(t) + B_2 \xi(t) && (59)\\ y(t) &= C_2 z_2(t) + \xi(t) && (60) \end{aligned}$$

which we can rewrite as

$$\begin{aligned} z_1(t+1) &= A_1 z_1(t) + B_1 x(t) && (61)\\ z_2(t+1) &= A_2 z_2(t) + B_2 C_1 z_1(t) + B_2 x(t) && (62)\\ y(t) &= C_2 z_2(t) + C_1 z_1(t) + x(t) && (63) \end{aligned}$$

$$\begin{pmatrix} z_1(t+1) \\ z_2(t+1) \end{pmatrix} = \begin{pmatrix} A_1 & 0 \\ B_2 C_1 & A_2 \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} x(t) \quad (64)$$

$$y(t) = \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + x(t). \quad (65)$$

It is instructive to compute the composition of a system with its inverse:

$$\begin{pmatrix} z_1(t+1) \\ z_2(t+1) \end{pmatrix} = \begin{pmatrix} A & 0 \\ BC & A - BC \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + \begin{pmatrix} B \\ B \end{pmatrix} x(t) \quad (66)$$

$$y(t) = \begin{pmatrix} C & -C \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + x(t). \quad (67)$$

This does not look exactly like $y(t) = x(t)$ yet. However the controllability Krylov matrix is a clue

$$\begin{pmatrix} \begin{pmatrix} B \\ B \end{pmatrix} & \begin{pmatrix} A & 0 \\ BC & A - BC \end{pmatrix} \begin{pmatrix} B \\ B \end{pmatrix} & \begin{pmatrix} A & 0 \\ BC & A - BC \end{pmatrix}^2 \begin{pmatrix} B \\ B \end{pmatrix} & \cdots \end{pmatrix} = \begin{pmatrix} B & AB & A^2 B & \cdots \\ B & AB & A^2 B & \cdots \end{pmatrix} \quad (68)$$

which is rank deficient, so the system is not minimal. The removal of the redundant rows involves adding the corresponding columns of the observability Krylov matrix, which annihilates them, showing that the Hankel matrix is zero. More directly, one can see that

$$\begin{aligned} \begin{pmatrix} C & -C \end{pmatrix} \begin{pmatrix} A & 0 \\ BC & A - BC \end{pmatrix}^k \begin{pmatrix} B \\ B \end{pmatrix} &= \begin{pmatrix} C & -C \end{pmatrix} \begin{pmatrix} A^k B \\ A^k B \end{pmatrix} \\ &= 0. && (69) \end{aligned}$$

So the composition of a system with its inverse does end up with the identity convolution; but the state space parameters of the composition are not minimal, obscuring the triviality of the system.

This example also makes clear that for any given impulse response, there are many sets of nonminimal state space parameters which correspond to that impulse response. This has important implications for system identification, as we will see (much later). But the situation is even worse, as we will see (immediately).

6. Change of State Space Coordinates

As usual, we start from the state space system

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) && (70)\\ y(t) &= Cz(t) + x(t) && (71) \end{aligned}$$

and now we observe that we can change the state space coordinates $z \mapsto Tz$, where T is any nonsingular transformation:

$$\begin{aligned} Tz(t+1) &= \left(TAT^{-1}\right) Tz(t) + (TB)\, x(t) && (72) \\ y(t) &= \left(CT^{-1}\right) Tz(t) + x(t) && (73) \end{aligned}$$

corresponding to the change of parameters

$$(A, B, C) \mapsto \left(TAT^{-1}, TB, CT^{-1}\right) \tag{74}$$

and the impulse response (and Hankel matrix) is preserved

$$\begin{aligned} \left(CT^{-1}\right)\left(TAT^{-1}\right)^k (TB) &= CT^{-1}TA^kT^{-1}TB && (75) \\ &= CA^kB. && (76) \end{aligned}$$

In the classical introduction to state space, this fact is sometimes referred to as a license to choose any coordinate system whatsoever for state space, since the impulse response (and the 'input output' relationship between x and y) is unaffected. It is a good idea to remember that the impulse response is invariant under change of state space coordinates, but it turns out that most coordinate systems for state space are not very good.

We can see that observability, controllability, and minimality are all preserved under change of state space coordinates, since

$$\begin{aligned} P &\mapsto TPT^* && (77) \\ Q &\mapsto T^{-*}QT^{-1} && (78) \end{aligned}$$

and this transformation preserves positive definiteness of P and Q (e.g. by Sylvester inertia).

What could go wrong? In fact it is a question of what could go right. But what occurs is that the change of state space coordinate does not preserve the eigenvalues of P and Q. We already know enough to worry if an eigenvalue of P or Q is very small compared to the others, since that is a system which may be minimal, but it is in some sense close to nonminimal. This is the first example of a geometric fact arising which can be in conflict with topology. We will have a lot more to say about this later.

However here it is time to point out that there are some eigenvalues which are preserved by the change of state space coordinates.

$$PQ \mapsto TPQT^{-1}. \tag{79}$$

The eigenvalues of PQ are invariant under state space coordinates, so by rights, these eigenvalues should have information about the impulse response. The relationship is fairly obvious in view of the factorization of the Hankel matrix

$$H = \begin{pmatrix} C \\ CA \\ \vdots \end{pmatrix} \begin{pmatrix} B & AB & \cdots \end{pmatrix} \tag{80}$$

which extends to a factorization of the Hermitian square

$$HH^* = \begin{pmatrix} C \\ CA \\ \vdots \end{pmatrix} \begin{pmatrix} B & AB & \cdots \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ \vdots \end{pmatrix} \begin{pmatrix} C^* & A^*C^* & \cdots \end{pmatrix}. \tag{81}$$

The eigenvalues of this matrix are the squares of the singular values of H. The relevant fact of linear algebra here is that the nonzero eigenvalues of a product of two matrices are the same if the two matrices are permuted, so the squares of the nonzero singular values of H are the eigenvalues of the matrix

$$\begin{pmatrix} B & AB & \cdots \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ \vdots \end{pmatrix} \begin{pmatrix} C^* & A^*C^* & \cdots \end{pmatrix} \begin{pmatrix} C \\ CA \\ \vdots \end{pmatrix} = PQ. \tag{82}$$

The nonzero singular values of H inform us as to the rank of a minimal factorization, but more importantly, the small singular values suggest whether or not H is close to a lower rank impulse response.

7. State Space as a Data Compression

The state space parameters provide a factorization of the Hankel matrix of the system, and that factorization neatly segregates into the past

$$z(t) = \begin{pmatrix} B & AB & \cdots \end{pmatrix} \begin{pmatrix} x(t-1) \\ x(t-2) \\ \vdots \end{pmatrix} \tag{83}$$

and future

$$\begin{pmatrix} E(y(t)|t-1) \\ E(y(t+1)|t-1) \\ \vdots \end{pmatrix} = \begin{pmatrix} C \\ CA \\ \vdots \end{pmatrix} z(t). \tag{84}$$

The infinite dimensional space of past innovations is the direct sum of the row space of the controllability Krylov matrix $\begin{pmatrix} B & AB & \cdots \end{pmatrix}$ and it's orthogonal complement. The component of the innovations sequence perpendicular to this row space is permanently *forgotten*, as far as the state space system is concerned, the only aspect of history that the state space system uses is that part which is in the row space of the controllability Krylov matrix. The components of the state $z(t)$ are coefficients in the expansion of the history of the innovations in terms of the 'functions' given by the rows of the the controllability Krylov matrix. Likewise, the components of the state are coefficients in the expansion of the conditional expectations of the future values of y in terms of the columns of the the observability Krylov matrix. The main subtlety to remember in this point of view is that the two expansions are related by the common A.

Suppose for example imagined a system with

$$z(t) = \begin{pmatrix} 1 & \lambda & \lambda^2 & \cdots \end{pmatrix} \begin{pmatrix} x(t-1) \\ x(t-2) \\ x(t-3) \\ \vdots \end{pmatrix} \tag{85}$$

and future

$$\begin{pmatrix} E(y(t)|t-1) \\ E(y(t+1)|t-1) \\ E(y(t+2)|t-1) \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ \mu \\ \mu^2 \\ \vdots \end{pmatrix} z(t). \tag{86}$$

This turns out be a linear time invariant system only if $\lambda = \mu$, otherwise

$$\begin{pmatrix} 1 \\ \mu \\ \mu^2 \\ \vdots \end{pmatrix} \begin{pmatrix} 1 & \lambda & \lambda^2 & \cdots \end{pmatrix} \tag{87}$$

is not Hankel.

8. Minimal Representations with the Same Response

We have already seen that two different state space systems can have the same response:

$$z(t+1) = 0z(t) + 1x(t) \tag{88}$$
$$y(t) = 0z(t) + x(t) \tag{89}$$

and

$$\begin{pmatrix} z_1(t+1) \\ z_2(t+1) \end{pmatrix} = \begin{pmatrix} A & 0 \\ BC & A - BC \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + \begin{pmatrix} B \\ B \end{pmatrix} x(t) \tag{90}$$
$$y(t) = \begin{pmatrix} C & -C \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} + x(t). \tag{91}$$

However these parameterizations correspond to different rank factorizations of the same Hankel matrix. If there are two different factorizations of the Hankel matrix with minimal rank, then the resulting state space systems differ only by a change of state space coordinates. Suppose that

$$H = \begin{pmatrix} C_1 \\ C_1A_1 \\ C_1A_1^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B_1 & A_1B_1 & A_1^2B_1 & \cdots \end{pmatrix} \tag{92}$$
$$= \begin{pmatrix} C_2 \\ C_2A_2 \\ C_2A_2^2 \\ \vdots \end{pmatrix} \begin{pmatrix} B_2 & A_2B_2 & A_2^2B_2 & \cdots \end{pmatrix} \tag{93}$$

are two factorizations with full rank. Then putting

$$T = \begin{pmatrix} C_1 \\ C_1A_1 \\ C_1A_1^2 \\ \vdots \end{pmatrix}^+ \begin{pmatrix} C_2 \\ C_2A_2 \\ C_2A_2^2 \\ \vdots \end{pmatrix} \tag{94}$$

we have

$$\begin{pmatrix} B_1 & A_1B_1 & A_1^2B_1 & \cdots \end{pmatrix} = T \begin{pmatrix} B_2 & A_2B_2 & A_2^2B_2 & \cdots \end{pmatrix} \tag{95}$$
$$= \begin{pmatrix} (TB_2) & \left(TA_2T^{-1}\right)(TB_2) & \left(TA_2T^{-1}\right)^2(TB_2) & \cdot \end{pmatrix} \tag{96}$$

so

$$B_1 = TB_2 \tag{97}$$

and

$$A_1 = \begin{pmatrix} AB_1 & A_1^2B_1 & A_1^3B_1 & \cdots \end{pmatrix} \begin{pmatrix} B_1 & A_1B_1 & A_1^2B_1 & \cdots \end{pmatrix}^+ = TA_2T^{-1} \tag{98}$$

from

$$\begin{aligned} C_1 \begin{pmatrix} B_1 & A_1B_1 & A_1^2B_1 & \cdots \end{pmatrix} &= C_1T \begin{pmatrix} B_2 & A_2B_2 & A_2^2B_2 & \cdots \end{pmatrix} && (99) \\ &= C_2 \begin{pmatrix} B_2 & A_2B_2 & A_2^2B_2 & \cdots \end{pmatrix} && (100) \end{aligned}$$

we finally have

$$C_1 = C_2T^{-1}. \tag{101}$$

There are many implications of this characterization. The Kalman decomposition is the fact that there are four invariant spaces of A:

$$\operatorname{row} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} \cap \operatorname{col} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \text{ controllable and observable space} \tag{102}$$

$$\operatorname{row} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} \cap \operatorname{col} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}^C \text{ uncontrollable but observable space} \tag{103}$$

$$\operatorname{row} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix}^C \cap \operatorname{col} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \text{ controllable but unobservable space} \tag{104}$$

$$\operatorname{row} \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix}^C \cap \operatorname{col} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}^C \text{ uncontrollable and unobservable space.} \tag{105}$$

It is clear that these are invariant spaces of A; for example if $z \in \operatorname{col} \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}$ then there is x such that

$$z = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} x \tag{106}$$

and so

$$Az = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} Zx \tag{107}$$

where Z is the (lower) shift. These invariant spaces may or may not be empty, for example if (A, B, C) are minimal, then only the controllable and observable space is nonempty.

9. TRANSFER FUNCTIONS

10. ONE VIEW OF PERTURBATION OF LINEAR SYSTEMS

If we perturb $z(t)$ to $\zeta(t) = z(t) \mapsto (\mathbb{I} + E) z(t)$ then the difference in the mapping from x to $\hat{y}$ is given by

$$\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} E \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}. \tag{108}$$

We then have

$$\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} E \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} = U_Q Q^{1/2} E P^{1/2} U_P^* \tag{109}$$

and in a unitarily invariant norm we have

$$\left\| \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \end{pmatrix} E \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \right\| = \left\| P^{1/2} E Q^{1/2} \right\| \tag{110}$$

so a measure of the sensitivity of the map to perturbations in $z(t)$ is $\left\| Q^{1/2} E P^{1/2} \right\|$ and a measure of the relative sensitivity is

$$\frac{\left\| Q^{1/2} E P^{1/2} \right\|}{\left\| Q^{1/2} P^{1/2} \right\|} \tag{111}$$

(we can interpret this as a 'noise to signal ratio'). It is important to note that the relative sensitivity depends on the state space coordinates; therefore the sensitivity depends on the implementation of the filter, not the filter itself (i.e. impulse response, transfer function). It follows that as much as possible, one would like to choose the state space coordinates appropriately.

For unitarily invariant norms we have the estimates for the perturbation

$$\begin{aligned} \left\| Q^{1/2} E P^{1/2} \right\| &\leq \left\| \Sigma_Q \Sigma_E \Sigma_P \right\| && (112) \\ &\leq \left\| \Sigma_Q \Sigma_P \right\| \sigma_1(E). && (113) \end{aligned}$$

where $Q^{1/2} = V_Q \Sigma_Q V_Q^*$, $P^{1/2} = V_P \Sigma_P V_P^*$ and $E = U_E \Sigma_E V_E^*$, (and the last inequality follows from the monotonicity of symmetric gauge functions). Then the relative error in the map is bounded by

$$\frac{\left\| Q^{1/2} E P^{1/2} \right\|}{\left\| Q^{1/2} P^{1/2} \right\|} \leq \frac{\left\| \Sigma_Q \Sigma_P \right\|}{\left\| Q^{1/2} P^{1/2} \right\|} \sigma_1(E) \tag{114}$$

and this bound has a factor $\sigma_1(E)$ which depends only on the perturbation and a factor $\frac{\left\| \Sigma_Q \Sigma_P \right\|}{\left\| Q^{1/2} P^{1/2} \right\|}$ which depends only on the system Grammians.

Writing $Q^{1/2}P^{1/2} = V_Q \Sigma_Q V_Q^* V_P \Sigma_P V_P^*$ we have

$$\left\| Q^{1/2}P^{1/2} \right\| = \left\| \Sigma_Q V_Q^* V_P \Sigma_P \right\| \geq \left\| \Sigma_Q J \Sigma_P \right\| \tag{115}$$

that is:

$$\frac{\left\| Q^{1/2} E P^{1/2} \right\|}{\left\| Q^{1/2}P^{1/2} \right\|} \leq \frac{\left\| \Sigma_Q \Sigma_P \right\|}{\left\| \Sigma_Q J \Sigma_P \right\|} \sigma_1 (E) . \tag{116}$$

This estimate provides a measure of how important the state space coordinate choice is: it appears only important when

$$\frac{\left\| \Sigma_Q \Sigma_P \right\|}{\left\| \Sigma_Q J \Sigma_P \right\|} \tag{117}$$

is large. In particular, if we choose $\Sigma_P = I$ or $\Sigma_Q = I$ then we have safely avoided this sort of sensitivity.

10.1. All Pass Perturbation. We give an example of perturbation sensitivity when the state space coordinate is chosen badly. It is one of a wide array of bad choices, but it is good to start somewhere.

Let one realization of a filter be

$$\begin{aligned} z(t+1) &= \Lambda z(t) + \beta x(t) && (118) \\ y(t) &= \gamma^T z(t) + \delta x(t) && (119) \end{aligned}$$

where Λ is diagonal and β and γ are vectors. Then in response to the monochromatic input $x(t) = e^{i\omega t}$ (with real frequency ω) the state components satisfy

$$z_k(t+1) = \lambda_k z_k(t) + \beta_k e^{i\omega t} \tag{120}$$

which we solve as

$$\begin{aligned} z_k(t+1) &= \beta_k e^{i\omega t} + \beta_k \lambda_k e^{i\omega(t-1)} + \beta_k \lambda_k^2 e^{i\omega(t-2)} + \cdots && (121) \\ &= \beta_k e^{i\omega t} \left(1 + \lambda_k e^{-i\omega} + \lambda_k^2 e^{-i2\omega} \right) && (122) \\ &= \frac{\beta_k e^{i\omega t}}{1 - e^{-i\omega}\lambda_k} && (123) \end{aligned}$$

and so the output satisfies

$$\begin{aligned} y(t) &= e^{i\omega t} \left(\delta + \sum_{k=1}^{n} \frac{\gamma_k \beta_k e^{-i\omega}}{1 - e^{-i\omega}\lambda_k} \right) && (124) \\ &= e^{i\omega t} \left(\delta + \sum_{k=1}^{n} \frac{\gamma_k \beta_k}{e^{i\omega} - \lambda_k} \right) . && (125) \end{aligned}$$

(The ouput is indeed the input $e^{i\omega t}$ multiplied by the transfer function evaluated at that frequency.)

However now suppose that we have chosen Λ, β, γ, and δ such that the filter is all-pass. That the filter is all-pass is that

$$\left| \delta + \sum_{k=1}^{n} \frac{\gamma_k \beta_k}{e^{i\omega} - \lambda_k} \right| = K \tag{126}$$

for all frequencies ω. It is also the same as that, putting $z = e^{i\omega}$ we have

$$\delta + \sum_{j=1}^{n} \frac{\gamma_j \beta_j}{z - \lambda_j} = K e^{i\theta} \prod_{j=1}^{n} \left(\frac{1 - \lambda_j^* z}{z - \lambda_j} \right) \tag{127}$$

which allows us to compute $\gamma_k \beta_k$ by equating coefficients in the Laurent expansion around $z = \lambda_k$.

$$\lim_{z \to \lambda_k} \left(\left(\frac{1}{z - \lambda_k} \right) \left(\delta + \sum_{j=1}^{n} \frac{\gamma_j \beta_j}{z - \lambda_j} \right) \right) = \lim_{z \to \lambda_k} \left(\left(\frac{1}{z - \lambda_k} \right) K e^{i\theta} \prod_{j=1}^{n} \left(\frac{1 - \lambda_j^* z}{z - \lambda_j} \right) \right) \tag{128}$$

$$\gamma_k \beta_k = K e^{i\theta} \left(1 - |\lambda_k|^2 \right) \prod_{j \neq k} \left(\frac{1 - \lambda_j^* \lambda_k}{\lambda_k - \lambda_j} \right) \tag{129}$$

thus

$$|\gamma_k \beta_k| = K \left(1 - |\lambda_k|^2 \right) \prod_{j \neq k} \left| \frac{1 - \lambda_j^* \lambda_k}{\lambda_k - \lambda_j} \right|. \tag{130}$$

For the filter to be stable requires $|\lambda_k| < 1$ for all k. This means that

$$\left| \frac{1 - \lambda_j^* \lambda_k}{\lambda_k - \lambda_j} \right| > 1. \tag{131}$$

In the case where there is a stability margin, that is $|\lambda_k| \leq \rho < 1$ for some ρ, then we have

$$\left| \frac{1 - \lambda_j^* \lambda_k}{\lambda_k - \lambda_j} \right| \geq \frac{1 - \rho^2}{2\rho} \tag{132}$$

and this provides

$$|\gamma_k \beta_k| \geq K \left(1 - |\lambda_k|^2 \right) \left(\frac{1 - \rho^2}{2\rho} \right)^{n-1} \tag{133}$$

$$\geq 2K \frac{\left(1 - \rho^2 \right)^{n-1}}{(2\rho)^n} \tag{134}$$

so that the values of $\frac{|\gamma_k \beta_k|}{K}$ are exponentially large in n.

But in the transfer function

$$\left| \delta + \sum_{k=1}^{n} \frac{\gamma_k \beta_k}{e^{i\omega} - \lambda_k} \right| = K \tag{135}$$

we see that the denominators satisfy $\left| e^{i\omega} - \lambda_k \right| < 2$, which is not large enough to change the exponential growth of $|\gamma_k \beta_k|$ so that

$$K = \left| \delta + \sum_{k=1}^{n} \frac{\gamma_k \beta_k}{e^{i\omega} - \lambda_k} \right| \ll |\delta| + \sum_{k=1}^{n} \left| \frac{\gamma_k \beta_k}{e^{i\omega} - \lambda_k} \right| \tag{136}$$

that is, the transfer function expression is dominated by cancellation and the concommitant potential for error.

We can see a little more is true if we consider that the factor

$$\left| \frac{1 - \lambda_j^* \lambda_k}{\lambda_k - \lambda_j} \right| \tag{137}$$

is made larger and larger as $|\lambda_k - \lambda_j|$ decreases. Thus we expect clustered poles to be worse for this realization of the filter than poles well separated in the unit disc in the hyperbolic metric.

So it seems that this realization is a dangerous way to compute this linear map. Because we can write the map explicitly as

$$\begin{aligned} y(t) &= Dx(t) + Cz(t) \\ z(t) &= Bx(t-1) + ABx(t-2) + A^2Bx(t-3) + \cdots \end{aligned}$$

we see that the catastrophic cancellation can in principle be controlled by changing the coordinate system of the state space

$$z(t) \rightarrow Tz(t) \tag{138}$$
$$A \rightarrow TAT^{-1} \tag{139}$$
$$B \rightarrow TB \tag{140}$$
$$C \rightarrow CT^{-1} \tag{141}$$
$$D \rightarrow D \tag{142}$$

In fact this is clear because for the allpass filter we have $PQ = \mu\mathbb{I}$ for some $\mu > 0$ and the change of state space coordinates acts as $PQ \rightarrow TPQT^{-1} = \mu\mathbb{I}$ then we can choose $T = Q^{1/2}$ so that $Q \rightarrow T^{-*}QT^{-1} = \mathbb{I}$ and $P \rightarrow \mu\mathbb{I}$. In this coordinate system, we have

$$\frac{\|\Sigma_Q \Sigma_P\|}{\|\Sigma_Q J \Sigma_P\|} = 1 \tag{143}$$

so there is no possible perturbation sensitivity.

Previous approaches in the literature for choosing coordinate systems are based on optimizing some measure. And in this particular case, the coordinate system where $P = \mu\mathbb{I}$ and $Q = \mathbb{I}$ coincides with (at least) three of these choices - it is input normal, output normal, and a balanced realizations.

On the other hand, we are interested in this particular approach because in practice it is not a problem if the sensitivity only satisfies a much weaker bound, say

$$\frac{\|\Sigma_Q \Sigma_P\|}{\left\|Q^{1/2}P^{1/2}\right\|} = \frac{\|\Sigma_Q \Sigma_P\|}{\left\|\Sigma_Q V_Q^* V_P \Sigma_P\right\|} \leq 100 \tag{144}$$

One of our objectives is to identify coordinate systems that provide this weaker perturbation sensitivity; by relaxing the perturbation sensitivity we expect to have more freedom to sastisfy other constraints on the system.

One interpretation of this would be that we want to characterize the 'level sets' of the function

$$\lambda\left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right) \tag{145}$$

where W is unitary and λ is the vector of eigenvalues. Note that $\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q$ is diagonal if and only if W is a permutation.

Conjecture 2. *The image of the map* $W \mapsto \log \lambda \left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right)$ *is a convex polyhedron with extreme points occurring where* W *is a permutation.*

This is a similar situation to the Schur-Horn theorem, and numerical computations verify this situation. This conjecture is the same as the conjecture that the indicated map is a moment map, allowing the application of the Kostant-Atiyah-Guilleman-Sternberg generalization of the Schur-Horn theorem. This conjecture is proved for 2×2 matrices. The logarithm is necessary in the conjecture because $\left|\det \Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right|$ is independent of W; thus $\Pi_k \lambda_k \left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right) = \left|\det \Sigma_P^2 \Sigma_Q^2\right|$, so the image of $\lambda \left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right)$ cannot be a convex set. On the other hand $\sum_k \log \lambda_k \left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right) = \log \left|\det \Sigma_P^2 \Sigma_Q^2\right|$ constrains the image of $\log \lambda \left(\Sigma_Q W \Sigma_P^2 W^* \Sigma_Q\right)$ to a hyperplane, which can support a convex set.