Hidden Input Observability

March 20, 2018

Abstract

We define the problem of dynamic systems with structural errors and explain how an auxiliary system with hidden inputs can help to determine such structural errors. For the case of linear systems, sufficient and necessary conditions in the means analytical properties are provided to check whether hidden inputs can be uniquely determined or not. We will call such systems hidden input observable. Furthermore we show how dynamic systems can be interpreted as networks and deduce a graphical criterion for hidden input observability. Finally we give examples for observable and unobservable systems and try to compute hidden inputs via the greedy-approach gradient-method.

Contents

1 Dynamic Systems **2**

2 Analytical Hidden Input Observability **6**
2.1 Sufficient Conditions 10
2.2 Necessary Condition 13
2.3 HIO Theorem 15

3 HIO as graphical criterion **17**
3.1 Basics of Graph Theory 18
3.2 Structural Dynamic Networks 21

1 Dynamic Systems

Whether you work in the fields of biology, chemistry or physics, or you deal with problems of applied sciences, the system you want to control, construct or investigate will probably be governed by underlying principles and laws of nature. These principles, either fundamental empirical, are usually formulated in means of *differential equations*. For example the famous Lotka-Volterra equations

$$\frac{\mathrm{d}}{\mathrm{d}t} N_1 = \alpha_1 N_1 - \beta_2 N_1 N_2 \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} N_2 = \beta_1 N_1 N_2 - \alpha_2 N_2 \quad , \tag{2}$$

that describe evolution of the number of preys N_1 and the number of predators N_2 in time, with some parameters that encode properties of the animals and the environment. Also the wave equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} - \frac{\mathrm{d}^2}{\mathrm{d}x^2} - \frac{\mathrm{d}^2}{\mathrm{d}y^2} - \frac{\mathrm{d}^2}{\mathrm{d}z^2} \right) \psi = 0 \quad , \tag{3}$$

that provides a mathematical equation for e.g. electromagnetic and also mechanical waves is a (partial) differential equation. An interesting class of differential equations is called *first order autonomous ordinary differential equations*. This kind of differential equations describes the time evolution of a vector x, called *state vector*, a numerical quantity that has the full information about the state of the considered system at each point in time. It takes the general form

$$\frac{\mathrm{d}}{\mathrm{d}t} x = f(x, u) \tag{4}$$

where x is an possibly n-dimensional vector, and u is a known external input called *control*. The function f encodes all the information about the system, that is all interactions within the system as well as the influence of the control. Therefore equation (4) is called the *system equation* and the system of interest is called a *dynamic system*. You have already seen the Lotka-Volterra equations as an example of 2-dimensional first order autonomous differential equation without external inputs.

Structural Error Estimation in Dynamical Systems It is the rule, rather than the exception, that we do not know the systems equation of a system we want to investigate. Most of the time we have a clue, though, how it could look like. This guess we will call the *nominal model* $\tilde{f}$, and the expected evolution of the state vector $\tilde{x}$. Let us assume that there principally exists a true f that governs the behaviour of the system, and for simplicity we ignore external inputs. Then

$$w := f(x) - \tilde{f}(x) \tag{5}$$

is a quantitative, time-resolved measure, whether the nominal and the true model coincide, or not. We call this *systematic model error* or simply *hidden input.*

One main problem in the sciences of dynamic systems is, that we have only limited knowledge about the state x. E.g. you can measure weather conditions, temperature, wind speed, the amount of rain etc. in a specific region at a specific time, but mathematically you would have to get all these data everywhere at all times, to predict the weather of the next day. This leads to the concept of the *observation function* h that maps the state x of a dynamic system to the measured quantities $y := h(x)$, called *observables.* Consequently this makes the determination of the quantity w a nontrivial task.

Let us collect the possible errors, we have to keep in mind when working with dynamic systems

- We work with a nominal $\tilde{f}$, that differs from the true f.
- In addition to the deterministic f there could be random processes, ϵ that influence the states.
- The observation function could be wrong, namely systematic measurement errors, and could have additional random terms, called measurement noise.

Assume we have data y^{obs} and simulated data $\hat{x}$ on the basis of $\tilde{f} + \hat{w}$. If we could minimize

$$||y^{\text{obs}} - h(\hat{x})||^2 \tag{6}$$

with respect to a guessed hidden input $\hat{w}$, is this optimal hidden input $\hat{w}^{\text{min}}$ the information we need to learn the true system f on basis of the data y^{obs}?

This question is the initial point of the SEED-project.

The last months showed that for a variety of low and medium dimensional systems, the *Dynamic Elastic Net* with its two approaches, the Greedy-Approach Gradient-Method and the Bayesian DEN, showed that they are able to reconstruct structural errors in dynamic systems to a high precision and therefore may provide be a fruitful tool improve the mathematical models of dynamic systems.

Hidden Input Observability The questions you directly derive from the Dynamic Elastic Net approach are:

- Does a optimal hidden input w exist?
- Can we give an algorithm to calculate this optimal hidden input?
- Is the solution unique?
- Is the procedure stable against measurement errors and numerical errors?

The framework for the determination of hidden inputs is given by the *optimal control theory*, where in this case the hidden input is the control that shall be optimized for archive an optimal fit to the data, subject to the governing systems equations. Thus some ideas and algorithmic approaches from this field can be adapted.

As the result of the work so far, the Greedy-Approach Gradient-Method is provides fast and reliable deterministic algorithm to compute hidden inputs by comparing data with a given nominal model. However there seem to exist types of dynamic systems, that give remarkably good results, where some dynamical systems seem to hide their hidden inputs very effectively. So the obvious questions is: What are the conditions, to guarantee uniqueness of the optimal hidden inputs? We call this property *hidden input observability*. We will start with the investigation of linear systems and we will see, that it is possible to give sufficient and necessary conditions in means of the algebraic properties of the system as well as the graphical representation as a network graph.

As a starting point of the mathematical treatment consider a true general linear system

$$\dot{x} = Ax + Bu \quad , \quad y = Cx \tag{7}$$

and a nominal guess

$$\dot{\tilde{x}} = \tilde{A}\tilde{x} + Bu \quad , \quad y = Cx \quad . \tag{8}$$

At this point we assume that the external inputs Bu and the observation function $x \mapsto Cx$ are known. Now imagine we add two different hidden inputs w and w' to the system such that the solution $\hat{x}$

$$\dot{\hat{x}} = \tilde{A}\hat{x} + Bu + Dw \quad , \quad \hat{y} = C\hat{x} \tag{9}$$

coincides with the observations, and analogous does $\hat{y}' = C\hat{x}'$, the solution for w'. Due to linearity we can subtract $\hat{y}'$ from $\hat{y}$ to get

$$\dot{\delta x} = \tilde{A}\delta x + D\delta w \quad , \quad \delta y = C\delta x \tag{10}$$

where $\delta x = \hat{x} - \hat{x}'$ and analogous for w and y. We deduce

A linear system is hidden input observable,
if the equations (10) provide an injective.

2 Analytical Hidden Input Observability

We consider the mapping $L^2([0,T])^{\otimes m} \to L^2([0,T])^{\otimes p}$ defined by

$$y(t) := C \int_0^t \mathrm{e}^{A(t-\tau)} D w(\tau)\,\mathrm{d}\tau \tag{11}$$

where $w : [0,T] \to \mathbb{R}^m$, $y : [0,T] \to \mathbb{R}^p$, $A \in \mathbb{R}^{n\times n}$, $D \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$.

We assume that in biological systems the functions w and y are defined on a natural interval of time that covers $[0,T]$. This means, that for a small ϵ the model could be extended to an interval $(0-\epsilon, T+\epsilon)$ and thus differentiation of w and y at $t=0$ and $t=T$ makes sense. To denote this idea without introducing ϵ we use the notation $[0^-, T^+]$.

For any function $f(t,\tau)$ it is easy to see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^t f(t,\tau)\,\mathrm{d}\tau = \int_0^t \frac{\partial f}{\partial t}(t,\tau)\,\mathrm{d}\tau + f(t,t) \quad . \tag{12}$$

Using the latter equation we get the derivatives of y as

$$y^{(q+1)}(t) = CA^{q+1} \int_0^t \mathrm{e}^{A(t-\tau)} D w(\tau)\,\mathrm{d}\tau + \sum_{l=0}^{q} CA^l D w^{(q-l)}(t) \tag{13}$$

where $y^{(q+1)}$ denotes the $q+1$-st derivative and analogous for w.

Definition 1

We write

1. $M_q := \left[(CA)\,(CAD)\,\ldots\,(CA^qD)\right]$ and $V_q := \operatorname{kernel} M_q$.
2. $P : \mathbb{R}^{(q+1)m} \to \mathbb{R}^{qm}$ is the "projection" operator that cuts away the first m components of a vector.
3. $\mathfrak{O}_{l,d} := \bigcap_{q=0}^{l} P^q V_{d+q}$ and $\mathfrak{O}_d := \lim_{l\to\infty} \mathfrak{O}_{l,d}$.
4. $W^{(q,r)}(t) := \left[w^{(q)}(t), w^{(q-1)}(t), \ldots, w^{(r)}(t)\right]^{\mathrm{T}}$ as a column vector.
5. $W^{(q)} := W^{(q,0)}$.

Remark 1

- M_q is understood as a concatenation of $p \times m$ matrices. Therefore it is a $p \times (q+1)m$ matrix.
- Also note that $V_r \subseteq \mathbb{R}^{(r+1)m}$ and therefore also $\mathfrak{O}_{q,r} \in \mathbb{R}^{(r+1)m}$.
- From the definitions we directly get $P^r W^{(q)} = W^{(q-r)}$.
- From the definition we see that $\mathfrak{O}_{l+1,d} = \mathfrak{O}_{l,d} \cap P^{l+1} V_{d+l+1}$ hence

$$\mathfrak{O}_{l+1,d} \subseteq \mathfrak{O}_{l,d} \quad . \tag{14}$$

Remark 2

Note, that P does not fulfil the projector property $P^2 = P$, therefore the word "projection" is in quotation marks.

Let U and V be subspaces of $\mathbb{R}^{(l+1)m}$. Then

$$P(U \cap V) \subseteq PU \cap PV \quad . \tag{15}$$

Proof. It is sufficient to show that $\hat{v} \in P(U \cap V)$ implies $\hat{v} \in PU \cap PV$. Let $\hat{v} \in P(U \cap V)$. Then exists $v \in U \cap V$ such that $Pv = \hat{v}$. That means $v \in U$ and thus $Pv \in PU$ analogous $Pv \in PV$. Hence $Pv = \hat{v} \in PU \cap PV$. □

As a consequence we see

$$V_d \cap P\mathfrak{O}_{l,d+1} \subseteq \mathfrak{O}_{l,d} \quad . \tag{16}$$

Proof. Using the latter equation

$$P(U_1 \cap U_2 \cap \ldots) \subseteq PU_1 \cap P(U_2 \cap \ldots) \subseteq PU_1 \cap PU_2 \cap \ldots \tag{17}$$

applied to $\mathfrak{O}_{l,d+1}$ this means

$$P\mathfrak{O}_{l,d+1} = P\left(\bigcap_{q=0}^{l} P^q V_{d+1+q}\right) \subseteq \bigcap_{q=0}^{l} P^{q+1} V_{d+1+q} \tag{18}$$

and hence

$$V_d \cap P\mathfrak{O}_{l,d+1} \subseteq V_d \cap PV_{d+1} \cap P^2 V_{d+2} \cap P^3 V_{d+3} \cap \ldots \cap P^{l+1} V_{d+1+l} \quad . \tag{19}$$

Since the right hand side equals $\mathfrak{O}_{l+1,d}$ we get

$$V_d \cap P\mathfrak{O}_{l,d+1} \subseteq \mathfrak{O}_{l+1,d} \quad . \tag{20}$$

As seen in remark 1 $\mathfrak{O}_{l+1,d} \subseteq \mathfrak{O}_{l,d}$. □

Finally if U is a sub vector space, then PU is a sub vector space.

Proof. We know for a vector space U and real number a

$$Pu, Pv \in PU \Rightarrow u, v \in U \Rightarrow a(u+v) \in U \Rightarrow P(a(u+v)) \in PU \tag{21}$$

and note that P is linear. □

Remark 3

For a fixed r we define $a_k := \dim \mathfrak{O}_{k,r}$. Since $\mathfrak{O}_{k+1,r} \subseteq \mathfrak{O}_{k,r}$ we see that a_k is a monotonically decreasing, integer valued sequence with lower bound 0. For that reason the limes

$$\lim_{k\to\infty} a_k := a^* \tag{22}$$

exists, is greater or equal 0 and there is a $N \in \mathbb{N}_0$ such that $a_N = a^*$.

Proof. Assume such a limits does not exist. Then for each integer b there is a $N(b)$ such that $a_{N(b)} < b \forall N > N(b)$. Since we know that this cannot hold for $b = 0$ we know that $a^* \geq 0$ exists. Now let a^* be the limits and assume there is no N such that $a_N = a^*$. Then, since a_k is integer valued, $a_k > a^*$ implies $a_k \geq a^* + 1$ for all k. That means a_k does not converge to a^* thus a^* would not be the limit. □

Therefore we know that $\mathfrak{O}_d = \lim_{l\to\infty} \mathfrak{O}_{l,d}$ is well defined and reached within a finite number of steps.

Proof. By definition $\mathfrak{O}_{l,d}$ are intersections of sub vector spaces thus again are vector spaces. As we already know $\mathfrak{O}_{l+1,d} \subseteq \mathfrak{O}_{l,d}$ thus for a fixed d there is a sequence of at most dm vectors f_i such that

$$\mathfrak{O}_{l,d} = \operatorname{span}\{f_1, f_2, \ldots, f_{i(l)}\} \quad . \tag{23}$$

We already know that $\dim \mathfrak{O}_{l,d} = i(l)$ and that there is an integer $N < \infty$ such that $i(l) \to i(N)$ as l increases. Therefore

$$\mathfrak{O}_{l,d} = \operatorname{span}\{f_1, f_2, \ldots, f_{i(N)}\} \quad \forall l \geq N \quad . \tag{24}$$

□

Finally

$$V_d \cap P\mathfrak{O}_{d+1} \subseteq \mathfrak{O}_d \quad . \tag{25}$$

Proof. We already know $V_d \cap P\mathfrak{O}_{l,d+1} \subseteq \mathfrak{O}_{l,d}$ holds for all l. Now we can increase l finitely many times to get the equation of interest. □

2.1 Sufficient Conditions

Lemma 1

For a fixed d
$$\exists N \in \mathbb{N}_0 | \mathfrak{O}_{N,d} = \{0\} \quad \Leftrightarrow \quad \mathfrak{O}_d = \{0\} \quad . \tag{26}$$

Proof. Due to $\mathfrak{O}_{l+1,d} \subseteq \mathfrak{O}_{l,d}$ the $\Rightarrow$ direction is obvious. We already know $\mathfrak{O}_d$ is well defined and reached within a finite number of steps thus also the $\Leftarrow$ direction holds. □

Lemma 2

For all $d \in \mathbb{N}_0$ we find

$$y \equiv 0 \quad \Rightarrow W^{(d)}(0) \in \mathfrak{O}_d \quad . \tag{27}$$

Proof. If $y \equiv 0$ we know that also each derivative $y^{(q)}(t) = 0$ at each point in time. Evaluation of (13) at $t = 0$ yields

$$\sum_{l=0}^{q} CA^l D w^{q-l}(0) = 0 \quad \forall q \in \mathbb{N}_0 \tag{28}$$

which is equivalent to

$$M_q W^{(q)}(0) = 0 \quad \forall q \in \mathbb{N}_0 \tag{29}$$

thus

$$W^{(q)}(0) \in V_q \quad \forall q \in \mathbb{N}_0 \quad . \tag{30}$$

Using the projection r times leads to

$$W^{(q-r)}(0) \in P^r V_q \quad \forall 0 \leq r \leq q \in \mathbb{N}_0 \quad . \tag{31}$$

Now it is interesting to take $d = q - r$ constant to get

$$W^{(d)}(0) \in P^r V_{d+r} \quad \forall 0 \leq r, d \in \mathbb{N}_0 \quad . \tag{32}$$

That means for a fixed d

$$W^{(d)}(0) \in P^0 V_d \cap P^1 V_{d+1} \cap P^2 V_{d+2} \cap \ldots \tag{33}$$

hence

$$W^{(d)}(0) \in \mathfrak{O}_d \quad \forall d \in \mathbb{N}_0 \tag{34}$$

□

Lemma 3

If $y \equiv 0$ and $W^{(k-1)}(0) = 0$ then

$$W^{(d+k,k)}(0) \in \mathfrak{O}_d \quad \forall d \in \mathbb{N}_0 \quad . \tag{35}$$

Proof. We already know

$$M_{q+k} W^{(q+k)}(0) = M_{q+k} \left[W^{(q+k,k)}(0), W^{(k-1,0)}(0) \right]^{\mathrm{T}} = 0 \quad \forall q \in \mathbb{N}_0 \quad . \tag{36}$$

This can be simplified to

$$M_q W^{(q+k,k)}(0) = 0 \quad \forall q \in \mathbb{N}_0 \quad . \tag{37}$$

As in lemma 1 we know that $W^{(q+k,k)}(0) \in V_q$ for all q leads to

$$W^{(d+k,k)}(0) \in \mathfrak{O}_d \quad \forall d \in \mathbb{N}_0 \quad . \tag{38}$$

□

Proposition 1: Sufficient Condition

If there are integers d and K such that $\mathfrak{O}_{K,d} = \{0\}$, then

$$y \equiv 0 \quad \text{in a vicinity of } t = 0 \quad \Rightarrow w^{(q)}(0) = 0 \quad \forall \quad q \in \mathbb{N}_0 \quad . \tag{39}$$

Proof. Let $y \equiv 0$ and assume we have found integers d^* and N such that $\mathfrak{O}_{N,d^*} = \{0\}$. From the lemma 2 we know $\mathfrak{O}_{d^*} = \{0\}$. Thus

$$W^{(d^*)}(0) = 0 \quad . \tag{40}$$

Assume we found and integer r with $W^{(rd^*+r-1)}(0) = 0$ and note that we found this with $r = 1$. The latter lemma with $k = rd^* + r$ and $d = d^*$ reads

$$W^{(rd^*+r-1)}(0) = 0 \quad \Rightarrow \quad W^{((r+1)d^*+r, rd^*+r)}(0) \in \mathfrak{O}_d \tag{41}$$

and using $W^{((r+1)d^*+r)} = \left[W^{((r+1)d^*+r, rd^*+r)}, W^{(rd^*+r-1)}\right]^{\mathrm{T}}$ this yields

$$W^{(rd^*+r-1)}(0) = 0 \quad \Rightarrow \quad W^{((r+1)d^*+r)}(0) = 0 \tag{42}$$

which completes the inductive proof. □

Example 1

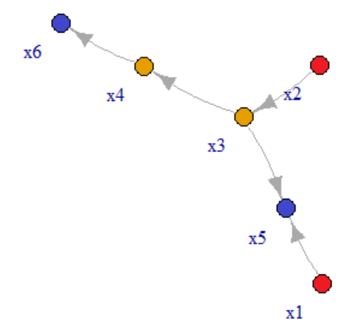
Consider the system depicted below, with hidden input knots x_1 and x_2 and observed knots x_5 and x_6. Writing e_i^r for the i-th canonical vector in $\mathbb{R}^r$. We see

$$V_1 = \text{span}\{e_1^4, e_2^4, e_4^4\} \tag{43}$$
$$V_2 = \text{span}\{e_1^6, e_2^6, e_3^6 - e_6^6, e_4^6, e_5^6\} \tag{44}$$
$$V_3 = \text{span}\{e_1^8, e_2^8, e_3^8 - e_6^8, e_4^8, e_5^8, e_7^8\} \tag{45}$$

and

$$PV_1 \cap P^2V_2 \cap P^3V_3 = \text{span}\{e_2^2\} \cap \text{span}\{e_2^2, e_1^2\} \cap \text{span}\{e_1^2\} = \{0\} \tag{46}$$

which means $\mathfrak{O}_{3,0} = \{0\}$. So this system fulfils the sufficient condition.



2.2 Necessary Condition

Definition 2

We define $\mathfrak{V}_0 := V_0 \backslash \{0\}$ and

$$\mathfrak{V}_i := V_i \cap P^{-1}\mathfrak{V}_{i-1} \quad . \tag{47}$$

Lemma 4

If and only if $\mathfrak{V}_d \neq \emptyset$ for all integers d, then it is possible to find a $\mathbb{R}^m$ valued sequence $(v_k)_{k\in\mathbb{N}_0}$ such that $v_0 \neq 0$ and

$$[v_k, v_{k-1}, \ldots, v_0] \in V_k \quad \forall k \in \mathbb{N}_0 \quad . \tag{48}$$

Proof. Assume for any integer K we find a sequence with $[v_k, \ldots, v_0]^\mathrm{T} \in V_k$ for all $k \leq K$ and $v_0 \neq 0$. Then we see by definition $v_0 \in \mathfrak{V}_0$. Now assume $[v_{k-1}, \ldots, v_0]^\mathrm{T}$ in $\mathfrak{V}_{k-1}$ and $[v_k, \ldots, v_0]^\mathrm{T}$ in V_k. Comparing with the definition shows that also $[v_k, \ldots, v_0]^\mathrm{T}$ in $\mathfrak{V}_k$. Thus with $v_0 \neq 0$

$$[v_k, \ldots, v_0] \in V_k \forall k \quad \Rightarrow \quad [v_k, \ldots, v_0] \in \mathfrak{V}_k \forall k \quad . \tag{49}$$

Now assume we found a $\mathfrak{V}_K = \emptyset$ and $\mathfrak{V}_k \neq \emptyset$ for all $k < K$. If there was a sequence of v_k such that $[v_k, \ldots, v_0]^\mathrm{T} \in V_k$ and $v_0 \neq 0$ for all $k < K$, then $[v_{K-1}, \ldots, v_0]^\mathrm{T} \in \mathfrak{V}_{d-1}$. By assumption there is no $\xi \in V_d$ with the property $P\xi \in \mathfrak{V}_{d-1}$. Thus there cannot be such a sequence. □

Proposition 2: Necessary Condition

If $\mathfrak{V}_d \neq \emptyset$ for all $d \in \mathbb{N}_0$, then we can find a nonzero function $w : I \to \mathbb{R}^m$ on an nonvanishing interval I such that $y \equiv 0$ on I.

Proof. By lemma 4 we know there is a sequence v_k such that $v_0 \neq 0$ and

$$[v_k, v_{k-1}, \ldots, v_0]^\mathrm{T} \in V_k \quad . \tag{50}$$

Thus if we define

$$w(t) := \sum_{l=0}^{\infty} \frac{t^l}{l!} v_l \tag{51}$$

we get $w^{(k)}(0) = v_k$ and $W^{(k)}(0) = [v_k, v_{k-1}, \ldots, v_0] \in V_k$. If we expand $y(t)$ around $t = 0$ we get

$$y(\tau) = y(0) + \sum_{r=0}^{\infty} \frac{\tau^{r+1}}{(r+1)!} y^{(r+1)}(0) = y(0) + \sum_{r=0}^{\infty} \frac{\tau^{r+1}}{(r+1)!} \underbrace{M_r W^{(r)}(0)}_{=0} \equiv 0 \tag{52}$$

as long as y can be expressed by its Taylor-expansion.

Note that since $v_0 \neq 0$, we know that $||[v_k, \dots, v_0]^{\mathrm{T}}||^2$ cannot converge to zero as k increases. $\square$

2.3 HIO Theorem

Proposition 3

To combine the necessary and sufficient condition, the implications

$$\mathfrak{V}_d \neq \emptyset \forall d \in \mathbb{N}_0 \quad \Longrightarrow \quad \mathfrak{D}_d \neq \{0\} \forall d \in \mathbb{N}_0 \tag{53}$$

equivalently

$$\exists K \in \mathbb{N}_0 \big| \mathfrak{D}_K = \{0\} \quad \Longrightarrow \quad \exists N \in \mathbb{N}_0 \big| \mathfrak{V}_N = \emptyset \tag{54}$$

hold.

Proof. First assume all $\mathfrak{V}_d$ are nonempty and we found a $\mathfrak{D}_N = \{0\}$. Then by the sufficient condition we know that there cannot be a function w such that $W^{(q)}(0) \in V_q$ for all q, but by the necessary condition we know that we can construct such a function.

The second implication is the complementary version of the first one. □

Theorem 1

If and only if we can find a $\mathfrak{V}_d = \emptyset$, then

$$y \equiv 0 \quad \Rightarrow \quad w^{(q)}(0) = 0 \quad \forall q \in \mathbb{N}_0 \quad . \tag{55}$$

If there are intervals I_i such that $\overline{\bigcup I_i} = [0^-, T^+]$ and such that w can be represented by its Taylor-Expansion on each interval then

$$\exists d \in \mathbb{N}_0 \,|\, \mathfrak{V}_d = \emptyset \quad \Longleftrightarrow \quad \text{HIO} \quad . \tag{56}$$

Proof. Proposition 2 already shows, that if we cannot find an empty $\mathfrak{V}_d$, then we can construct a function w such that $y \equiv 0$ on an interval, where its Taylor series converges. Thus the existence of such an empty $\mathfrak{V}_d$ is necessary for the implication.

Assume we found an empty $\mathfrak{V}_K$. By proposition 3 we know there is an $\mathfrak{D}_N = \{0\}$ which is sufficient to show that w must be a function such that $W^{(q)}(0) = 0$ for all q. Thus the existence of such a $\mathfrak{V}_K$ is also sufficient.

Now assume there are intervals as in the theorem. We write t_i for the supremum of I_i and without loss of generality we can assume that the infimum of I_{i+1} coincides with t_i. Then

$$w^{(q)}(0) = 0 \forall q \in \mathbb{N}_0 \quad \Leftrightarrow \quad w \equiv 0 \text{ on } I_1 \quad . \tag{57}$$

We now simply shrink the time domain to $[t_i^-, T^+]$ and straight forwardly get

$$w^{(q)}(t_i) = 0 \forall q \in \mathbb{N}_0 \quad \Leftrightarrow \quad w \equiv 0 \text{ on } I_{i+1} \tag{58}$$

which means w is the zero function on each interval. Therefore $w \equiv 0$ on $[0, T]$.
□

3 HIO as graphical criterion

Consider a dynamic system with matrices A, B, C and D with n state variables, p observables and m hidden inputs. As before we neglect the known inputs via B. Now assume each hidden input affects the state variables $x_{i_1}, x_{i_2}, \dots$ directly, that means the α-th column of D corresponds to w_α and can be represented by index set $\hat{D}_\alpha$ such that $a \in \hat{D}_\alpha$ means w_α hits x_a. The observable y_β corresponds to the β-th row of C. For that we can write $y_\beta = \sum_{b \in \hat{C}_\beta}$ for an index set $\hat{C}_\beta$.

Before we deal with graphical HIO we collect basics of graph theory together with definitions convenient for HIO.

3.1 Basics of Graph Theory

Definition 3: Graphtheoretical Definitions

1. When we consider a dynamic system as network, the state variables x_i shall be called *knots*.

2. Let A be the matrix of the linear dynamic system. The *path factor* of a sequence $\pi = (x_{l_0}, x_{l_1}, \ldots, x_{l_N})$ is

$$F(\pi) := \prod_{i=0}^{N-1} A_{l_{i+1} l_i} \quad . \tag{59}$$

3. We define a relation

$$x_a \sim x_b \quad \overset{\text{definition}}{\Longleftrightarrow} \quad F((x_a, x_b)) \neq 0 \quad \left(\quad \Longleftrightarrow \quad A_{ba} \neq 0 \quad \right) \quad . \tag{60}$$

4. A *path* from x_a to x_b is a sequence $(x_{l_0}, x_{l_1}, x_{l_2}, \ldots, x_{l_N})$ with indices l_i such that

$$x_{l_0} = x_a \quad , \quad x_{l_N} = x_b \quad \text{and} \quad x_{l_i} \sim x_{l_{i+1}} \quad . \tag{61}$$

5. The *length* of such a path is $||(x_{l_0}, \ldots, x_{l_N})|| = N$.

6. (a) For the *set of all paths* from x_a to x_b we write $\Gamma(x_a, x_b)$.
   (b) The *set of paths of length k* from x_a to x_b is $\Gamma_k(x_a, x_b)$.
   (c) For the *shortest path* from x_a to x_b (if it exists) we write $\gamma(x_a, x_b)$.

7. The *distance* between x_a and x_b is $d(x_a, x_b) := ||\gamma(x_a, x_b)||$ if it exists. If there is no path from x_a to x_b we write $d(x_a, x_b) = \infty$.

8. (a) The path (x_a) is called *trivial path*.
   (b) The path (x_a, x_a) is called *self-loop*.
   (c) Any path $(x_a, x_{l_1}, \ldots, x_{l_{N-1}}, x_a)$ is called *loop*.

9. Let π and π' be paths.
   (a) We write $x_i \in \pi$ if x_i appears in π.

Definition 4

Let $\pi = (x_{l_0}, \dots, x_{l_N})$ and $\pi' = (x_{k_0}, \dots, x_{k'_N})$ be sequences of knots.

1. $$\pi \odot \pi' := (x_{l_0}, \dots, x_{l_N}, x_{k_0}, \dots, x_{k'_N}) \tag{62}$$

2. If $x_{l_N} = x_{k_0}$.
$$\pi \oplus \pi' := (x_{l_0}, \dots, x_{l_N}, x_{k_1}, \dots, x_{k'_N}) \tag{63}$$

3. We write $\pi \subset \pi'$ if $\pi' = (x_{l_0}, \dots, x_{l_{N'}}) \oplus \pi \oplus (x_{l_{N'+||\pi||}}, \dots, x_{l_N})$.

Lemma 5

1. The path factor is a homomorphism, i.e.
$$F(\pi \oplus \pi') = F(\pi)F(\pi') \tag{64}$$
and
$$F(\pi \odot \pi') = F(\pi)F((x_{l_N}, x_{k_0}))F(\pi') \quad . \tag{65}$$

2. For a sequence $\pi = (x_{l_0}, \dots, x_{l_N})$
$$F(\pi) \neq 0 \quad \Leftrightarrow \quad \pi \text{ is a path} \quad . \tag{66}$$

Proof. 1. Write down the definitions to see that these equations hold.

2. Decompose $\pi = (x_{l_0}, x_{l_1}) \oplus (x_{l_1}, x_{l_2}) \oplus \dots$ to see that
$$F(\pi) = \prod_{i=0}^{N-1} F(x_{l_i}, x_{l_{i+1}}) = \prod_{i=0}^{N-1} A_{l_{i+1} l_i} \quad . \tag{67}$$
If and only if all factors $A_{l_{i+1} l_i} \neq 0$, then by definition π is a path.

□

Lemma 6

Let π be a path from x_a to x_b.

1. For two paths π and π', $||\pi|| + ||\pi'|| = ||\pi \oplus \pi'||$.
   (Addition of path lengths)

2. If any x_i appears more than once in π, then $\pi \neq \gamma(x_a, x_b)$.
   (Shortest paths contain no loops)

3. If there are $x_{l_i}, x_{l_j} \in \pi$ such that $x_{l_i} \sim x_{l_j}$ and $j \neq i + 1$, then $\pi \neq \gamma(x_a, x_b)$.
   (No shortcuts for shortest paths)

4. Let x_a, x_b ,$x_{a'}$ and $x_{b'}$ be knots. If π is a path from $x_{a'}$ to $x_{b'}$ and if $\pi \subset \gamma(x_a, x_b)$ then $\pi = \gamma(x_{a'}, x_{b'})$.
   (Subpaths of a shortest path are again a shortest path)

Proof.

1. Consider two paths $\pi = (x_{l_0}, \ldots, x_{l_N})$ and $\pi' = (x_{k_0}, \ldots, x_{k_{N'}})$ with the property $x_{l_N} = x_{k_0}$. Then $\pi \oplus \pi' = (x_{l_0}, \ldots, x_{l_N}, x_{l_{N+1}}, \ldots, x_{N+N'})$ with $l_{N+i} = k_i$. We see that $||\pi|| = N$, $||\pi'|| = N'$ and $||\pi \oplus \pi'|| = N + N'$.

2. Assume x_{l^*} appears twice in a path π from x_a to x_b. Then we can write $\pi = \pi_1 \oplus \pi^* \oplus \pi_2$ such that π_1 is a path from x_a to x_{l^*}, π^* is a nontrivial path from x_{l^*} to x_{l^*} and π_2 is a path from x_{l^*} to x_b. Thus also $\pi_1 \oplus \pi_2$ is a path from x_a to x_b and $||\pi_1 \oplus \pi_2|| = ||\pi|| - ||\pi^*||$ which means $\pi_1 \oplus \pi_2$ is shorter than π and therefore π cannot be the shortest path $\gamma(a, b)$.

3. We can write $\pi = \pi_1 \oplus \pi_2 \oplus \pi_3$ such that π_1 is from x_a to x_{l_i}, π_2 is from x_{l_i} to x_{l_j} and π_3 is from x_{l_j} to x_b. Since $x_{l_i} \sim x_{l_j}$ we can define $\pi^* = \pi_1 \oplus (x_{l_i}, x_{l_j}) \oplus \pi_3$. We see that $||\pi^*|| \leq ||\pi||$ and equality holds only if $||\pi_2|| = 1$ which means that $l_j = l_{i+1}$ or equivalently $j = i + 1$. Thus if $j \neq i + 1$ we have found a shorter path and π cannot be $\gamma(a, b)$.

4. We can write $\gamma(x_a, x_b) = \pi_1 \oplus \pi \oplus \pi_2$ where π_1 is from x_a to $x_{a'}$ and π_2 is from $x_{b'}$ to x_b. Now assume $\pi \neq \gamma(x_{a'}, x_{b'})$. Then $\pi^* = \pi_1 \oplus \gamma(x_{a'}, x_{b'}) \oplus \pi_2$ is shorter than $\gamma(x_a, x_b)$. This is in contradiction to the assumptions.

$\square$

3.2 Structural Dynamic Networks

Our aim is to get information about the matrices M_q and the sets V_q, so that we can apply the HIO theorem, just by properties of the influence graph.

Proposition 4

In the situation described above

$$(CA^kD)_{\beta\alpha} = F\left(\Gamma_k(w_\alpha, y_\beta)\right) \tag{68}$$

where the path factor F of a set is understood as the sum over the path factors of all elements and $\Gamma_k(w_\alpha, y\beta)$ is understood as the set of all paths in $\Gamma_k(x_a, x_b)$ where $a \in \hat{D}_\alpha$ and $b \in \hat{C}_\beta$.

Proof. At first note that

$$A^k_{ba} = \sum_{\pi\in\Gamma_k(x_a,x_b)} F(\pi) \quad . \tag{69}$$

This becomes obvious when we write

$$A^k_{ba} = \sum_{l_1,l_2,\dots,l_{k-1}=1}^{n} A_{bl_{k-1}} \dots A_{l_2,l_1} A_{l_1 a} \tag{70}$$

and note that the right hand side is just the sum over $F((x_a, x_{l_1}, \dots, x_b))$ with all possible values of $l_1, \dots, l_{k-1}$. As already seen $F(\pi) = 0$ is zero if π is not a path. Thus all sequences $(x_a, \dots, x_b)$ that give a nonzero path factor are paths of length k.

For simplicity we assume that w_α hits x_a. We find

$$(CA^kD)_{\beta\alpha} = \sum_{\lambda,\rho=1}^{n} C_{\beta\lambda} A^k_{\lambda\rho} D_{\rho\alpha} = \sum_{a\in\hat{D}_\alpha} \sum_{b\in\hat{C}_\beta} A^k_{ba} \quad . \tag{71}$$

By construction the α-th column of D is 1 if $\rho \in \hat{D}_\alpha$ and 0 else. The β-th row of C gives $C_{\beta\lambda} = 1$ if $\lambda \in \hat{C}_\beta$ and zero else.

Finally

$$(CA^kD)_{\beta\alpha} = \sum_{a\in\hat{D}_\alpha} \sum_{b\in\hat{C}_\beta} \sum_{\pi\in\Gamma_k(x_a,x_b)} F(\pi) \tag{72}$$

which is the right hand side of the proposition. □

Note that this proposition can easily be extended to the case of general matrices D and C by multiplying F with some factors. On the other hand we can simplify the proposition by using F as the indicator that gives 0 if the set is empty and 1 if not.

Definition 5

Define

$$K_i := \min_{\alpha\in\{1,\dots,m\}} d(w_\alpha, y_i) \tag{73}$$

i.e. the length of the shortest path from any knot that is affected by a hidden input to a knot that affects the observable y_i. Furthermore

$$\Xi := \begin{pmatrix} F\left(\Gamma_{K_1}(w_1,y_1)\right) & F\left(\Gamma_{K_1}(w_2,y_1)\right) & \dots \\ F\left(\Gamma_{K_2}(w_1,y_2)\right) & F\left(\Gamma_{K_2}(w_2,y_2)\right) & \dots \\ \vdots & \ddots & \ddots \end{pmatrix} . \tag{74}$$

If Ξ is injective, then the system is called *structural hidden input observable.*

Theorem 2

If a system is structural HIO, then it is HIO for almost all parameter settings.

Proof. Without loss of generality $K_1 \le K_2 \le \dots \le K_p$. We write the matrices M_k in rows for an appropriate k,

$$M_k = \left[\begin{pmatrix}0\\0\\0\\\vdots\end{pmatrix}\;\begin{pmatrix}\dots\\\dots\\\dots\\\vdots\end{pmatrix}\;\begin{pmatrix}(CA^{K_1}D)_{1.}\\0\\0\\\vdots\end{pmatrix}\;\begin{pmatrix}\dots\\\dots\\\dots\\\vdots\end{pmatrix}\;\begin{pmatrix}(CA^{K_2}D)_{1.}\\(CA^{K_2}D)_{2.}\\0\\\vdots\end{pmatrix}\;\begin{pmatrix}\dots\\\dots\\\dots\\\vdots\end{pmatrix}\;\begin{pmatrix}(CA^{K_3}D)_{1.}\\(CA^{K_3}D)_{2.}\\(CA^{K_3}D)_{3.}\\\vdots\end{pmatrix}\right] \tag{75}$$

We deduce

$$V_0 = \mathbb{R}^m \quad \text{and} \quad V_k = \mathbb{R}^{(k+1)m} \quad \text{for } k < K_1 \tag{76}$$

since M_k consists only of zeroes. Then from the first row we get

$$V_{K_1} = \mathbb{R}^{K_1 m} \times \left\{ \xi \in \mathbb{R}^m \middle| \sum_{l=1}^{m} \xi_l (CA^{K_1}D)_{1l} = 0 \right\} \tag{77}$$

and

$$P^{K_1} V_{K_1} = \left\{ \xi \in \mathbb{R}^m \middle| \sum_{l=1}^{m} \xi_l F\left(\Gamma_{K_1}(w_l, y_1)\right) = 0 \right\}. \tag{78}$$

From the second row we get

$$P^{K_2} V_{K_2} \subseteq \left\{ \xi \in \mathbb{R}^m \middle| \sum_{l=1}^{m} \xi_l F\left(\Gamma_{K_2}(w_l, y_2)\right) = 0 \right\} \tag{79}$$

and so on. Note that we neglect the information in the $(\dots)$ parts of equation (75), therefore the condition we get will be sufficient but not necessary.

Now we see that

$$\mathfrak{O}_0 \subseteq \bigcap_{r=1}^{p} P^{K_r} V_{K_r} \subseteq \bigcap_{r=1}^{p} \left\{ \xi \in \mathbb{R}^m \middle| \sum_{l=1}^{m} \xi_l F\left(\Gamma_{K_r}(w_l, y_r)\right) = 0 \right\} \tag{80}$$

and the right hand side of the latter equation is exactly

$$\left\{ \xi \in \mathbb{R}^m \middle| \sum_{l=1}^{m} \Xi \xi = 0 \right\} \quad . \tag{81}$$

Thus Ξ is injective leads to $\mathfrak{O}_0 = \{0\}$ which is sufficient for HIO. □

References

[1] Achim Ilchmann Thomas Berger. “Zero dynamics of time-varying linear systems”. In: (2010).